

Half-hearted war on drugs

The principle that for every important social problem, there should be a conference at which officials pledge their best efforts to its solution seems to have taken a tumble at last week's London conference on drugs.

THE British government, sponsor of last week's international conference on drugs in London, is probably disappointed that the event showed the participants to differ among themselves about the best strategy for dealing with this important problem. But it could not have been otherwise. Estimations of the best strategy differ chiefly because the problem is complex as well as important. Some countries, notably the United States, but also Britain itself, follow a similarly tough line on all narcotic drugs, but others (the Netherlands, for example) regard the use of cannabis as less serious than that of, say, heroin.

And while the sale and even use of narcotic drugs are illegal in most places, there are important differences in the mixture of penal sanctions and opportunities for treatment offered to those discovered to be addicts. That, again, is unavoidable, at least for the time being.

The difficulty is that there is a sharp distinction between short-term and long-term strategy. Given the rapid increase of addiction in the industrialized West to three successive waves of narcotic substances — cannabis, opium, now cocaine and their derivatives — it is natural that governments should direct their attention at the sources of supply of these substances and at the means by which they are distributed internationally as well as at measures calculated to reduce the demand for narcotics by the ultimate users. But hemp, poppies and cocoa leaves are not the only natural sources of narcotic alkaloids, while the LSD fashion of just a few years ago shows how much can be done in makeshift laboratories to meet local demands for addictive highs. In the long run, the only strategy that will banish the evil consequences of addiction is to encourage the reduction of demand.

Most governments are in a dilemma of their own making. Among the harmful substances known to be addictive, the most widely used are alcohol and tobacco, both of which cause great damage to people's health. In most industrialized countries, the need for treatment of these common addictions far exceeds the resources and skill likely ever to become available. Yet there are few countries in which the use of these substances is illegal. (The exceptions are some Moslem states, where the reasons are doctrinal, not medical.) That is as it should be. It would be a gross infringement of the civil rights of people who use alcohol in moderation have to give it up because some among them become addicts. (The case for continuing to allow the sale of tobacco is weaker, resting

chiefly on the difficulty of enforcement.) The logical difficulty is that governments appear to be prevented from learning what they might from alcohol and tobacco addiction by the fear of being discovered to be inconsistent.

That is a great misfortune, for there is much to learn from the ways in which addiction to alcohol and tobacco can be intelligently dealt with. Intensive treatment can wean individuals away from their addiction (as can be heroin addicts), but chic health farms are not a general recipe. Most governments tax both alcohol and tobacco, not simply from cupidity but so as to restrain demand by a market mechanism. Fiscal controls would not work with narcotic drugs (although cannabis may be in a different category): one of the greatest social evils of the wave of addiction in the West since the 1960s comes about from the way in which many addicts have to steal to acquire their supplies. But there are other than fiscal means of regulating demand — notably the provision of cheap supplies to those who register their addiction and who consent to continuing medical supervision.

If they are properly monitored, some future conference should learn much not only about the effectiveness of this arrangement but about the reasons that distinguish addicts from others, as well as of the social benefits of creating social links with a closed and self-isolating community.

Whatever the success of short-term measures against the sources of supply and the distribution networks, such devices must play an important part in the longer haul. But would they not condone illegality? Yes (unless, in the light of more and encouraging experience, the law were changed), but no more inconsistently than in the supply of sterile hypodermic needles to intravenous drug users to prevent the spread of HIV infection.

But there are also penal measures that may help. So much is clear from the way in which increasingly tough penalties for drunken driving are beginning to have an effect on the rate of serious road accidents. But a drug user is also the chief victim, with his or her family, meaning that imprisonment is an ineffective remedy.

What the war on drugs needs is a system of non-custodial sentences for drug users who break the laws which apply to them, preferably one that entails a large measure of the blend of exhortation and education without which young people will not be adequately warned of the dangers of being hooked. □



Embryo research

The British House of Commons should this week settle for the government's bill on embryo research.

THE bill to give effect to the recommendations of the Warnock committee, now five years stale, reaches the House of Commons this week, perhaps even today (see opposite). That is why it is to be hoped that all concerned will pay more than passing attention to the article on page 768 of this issue, which carries one of the first formal reports to have appeared in print of the potential benefits of the manipulation of human embryos before they have been implanted into the uterus.

What Professor Robert Winston and Dr Alan Handside have done is to exploit the property (well-known in, for example, the mouse) that not all the cells in an early embryo are necessary for its successful development. It is possible to take one cell from, say, the four-cell stage of the embryo's development, use that cell for making a diagnosis of the genetic condition of the embryo as a whole, and still look to the remaining cells to yield a viable fetus. What the work at the Hammersmith Hospital has shown is that it is possible to use this technique for screening embryos for their gender, using for implantation only female embryos, almost certain to be free from genetic defects linked with the X-chromosome, of which females carry two, one from each parent.

It is important and relevant to the Embryo Bill that it is entirely predicted that the technique would be effective: endless experiments with other mammals have given ample assurance that it would be. But this implies that the technique is an operational technique — a means of avoiding genetic defects that might otherwise be detectable only much later in pregnancy, perhaps resulting legitimately in abortion. It has been used, with approval of the voluntary committee set up to span the long wait for the Embryo Bill, and will no doubt be more generally applied when there is a legal framework in which other groups can operate. Defective embryos will, of course, be discarded. The ethical question for most people is whether the discarding of some out of several embryos at this stage is preferable to an abortion later. The House of Commons cannot be in much doubt where its decision should lie. In logic, those who support the present abortion law cannot easily oppose the Embryo Bill.

But that, of course, does not dispose of all the arguments. There are many who hold that neither abortion nor the manipulation of embryos should be allowed. The opinion, which has as a corollary the view that the birth of children with avoidable genetic defects is virtuous in some sense, deserves respect as a matter of civil comity. It is also relevant that, at least for recessive genes, there may be genetic arguments for caution: the disadvantages of the homozygote may be more easily recognized than the advantages of the heterozygote.

But that, fortunately, is but yet another argument for

caution, which is one of the considerations that should weigh with the Licensing Authority which, the British Government proposes, will replace the voluntary committee that has stood in for it. And, properly briefed, that authority should give the lie to those who also argue that the manipulation of embryos before implantation is the beginning of a dangerous slippery slope. □

Journalist's dilemma

The British press has been placed in an intolerable position by a court decision last week.

SUPPOSE a member of *Nature's* editorial staff were told on the telephone that a British computer manufacturer — call it Chipmunk Ltd — had run out of cash and was trying to raise money from its bankers. Suppose that the recipient of this information then called the company to establish whether what he had been told was true. The precedent of a bizarre court case last week suggests the following course of events might follow. First, Chipmunk would apply to a High Court judge for an injunction to prevent publication of the news on the grounds that it would damage its commercial interests. Then, arguing that the information could come only from confidential documents, Chipmunk would demand that the journalist involved should reveal the name of his informant, meanwhile winning the courts' approval for a further injunction that Chipmunk's name should not be mentioned in any reports of its action. Relying on a strict interpretation of the Contempt of Court Act, 1981, which says that journalists may withhold the identity of the sources of their information except when disclosure is in the "national interest or the interest of justice", the court would then order the journalist to hand over his notebook on pain of being found guilty of contempt of court.

This is almost exactly what happened to Mr William Goodwin, a young journalist on *The Engineer*, who was fined £5,000 last week for his 'offence'. But the judges made it plain that he should count himself lucky; he might have been sent to jail for two years instead.

The puzzle in this is that the concept of justice seems to have become deranged. The original intention of the act was that journalists' freedom not to name their sources would be attenuated only when criminal prosecutions would thereby be impeded (which is itself offensive). In this case, while Chipmunk may have had a perfectly understandable wish to trace the leak of information, there can be no issue of justice involved.

Even if Chipmunk had a case for believing it would be damaged by publication, all damage has been prevented by the injunction on publication even of its name. Apart from the scant concern shown both by the law and by the judges who have interpreted it for the freedom of expression in Britain, the lesson of the Goodwin case will weigh with all potential whistle-blowers. If this is what the law means, it should be changed. □

Pressure stepped up on embryo research

■ Close result expected in Commons vote

■ Undecided MPs heavily lobbied

London

As parliamentary lobbying to influence the House of Commons vote on human embryo research in Britain reaches a crescendo, the opposing sides agree on one thing alone — the result will be close.

On 23 April, Members of Parliament (MPs) vote on alternative versions of Clause 11 of the Human Fertilisation and Embryology Bill, free from an official party line. One of these alternatives follows the recommendations of the 1984 Warnock Report, allowing carefully regulated research up to 14 days from conception. The other would make this research a criminal offence. The pro-research lobby, led by PROGRESS, an umbrella organization of charities, medical and scientific groups, argues that future developments in infertility treatment, contraception and the diagnosis and treatment of genetic disease would be prevented by a ban on research.

In the House of Lords, peers voted in favour of research by nearly three to one (see *Nature* 343, 577; 15 February 1990). But Christine Lavery, secretary of the Genetics Interest Group (GIG), representing a number of genetic disease charities, says that peers were better informed than most MPs. Professor Robert Winston, of Hammersmith Hospital, London, and president of PROGRESS (whose paper on embryo research is published on page 768 of this issue), echoes this view: "If Members of Parliament were fully educated", he says, "I can't believe they would vote against research".

Both the Medical Research Council (MRC) and the Royal Society are taking the unusual step, for government-funded bodies, of contacting MPs to correct errors of scientific fact creeping into the debate. Professor Richard Gardner, chairman of the Royal Society's *ad hoc* committee on embryo research, is also worried by the assertion from many anti-research MPs and pressure groups that they are not against 'non-destructive research' (when embryos are placed back into the mother). Apart from scientific considerations, allowing embryos used in experimental procedures to go to full-term would be "grossly irresponsible", he says. With research at an early stage, the effects of some procedures on later development are not yet known.

For organizations involved in more overt lobbying of MPs, identifying those who have not yet made up their minds may be crucial, so that resources can be

focused in the run-up to the debate. Lavery says that GIG is concentrating on a "grey area" of about 75 MPs, while LIFE, a 'pro-life' group campaigning on the anti-research side, is targeting 250 MPs who did not attend the bill's second reading debate, earlier this month.

Both sides claim that opposing lobbyists have spent far more than themselves. Phyllis Bowman, of the Society for the Protection of Unborn Children (SPUC), largest group in the anti-research lobby, describes claims that her organization has channelled upwards of £250,000 into the campaign as "rubbish". SPUC's accounts are not yet available, she says, but adds that she has mortgaged her own home to help finance the campaign, and attacks the MRC's involvement as an abuse of public funds. Winston says that PROGRESS has been limited to a budget of £20,000 over three years.

The government is also under attack, for its handling of the bill. Peter Thurnham, a Conservative MP and member of PROGRESS, says that holding the debate on a Monday evening, shortly after the Easter recess, with many MPs still in their constituencies, may mean a low attendance. Although Thurnham is confident that a majority of MPs favour research, a low turnout could give a "maverick" result, he says.

Pro-research groups oppose the government's decision to allow a clause on abortion to be added to the bill (see *Nature* 344, 476; 5 April 1990). Deputy Prime Minister Sir Geoffrey Howe's clause proposes a cut in the upper time limit for abortion from 28 to 24 weeks of pregnancy, and other MPs have tabled amendments that propose even lower time limits. Although the abortion clause will be discussed the following day (24 April), the fear is that the emotionally charged arguments of the abortion debate may spill over into the discussion of the embryo research clause, and cloud what should be a separate issue.

Opinion among the anti-research lobby on this question is divided. Bowman from SPUC agrees that the two issues would be better dealt with separately, but Nigel Williams from Christian Action, Research and Education (CARE) and Conservative MP Ann Widcombe both welcome the inclusion of a clause on late abortion in a government-sponsored bill.

The abortion and embryo research issues are already being linked together by some members of the anti-research lobby.

A pamphlet issued to other MPs by Sir Bernard Braine, a Conservative member, in response to an earlier PROGRESS document, claims that research to identify pre-implantation tests to screen for genetic diseases will mean "aborting those who suffer from genetic disease". Widcombe also rejects the separation of the two issues: "Pro-life MPs believe that life begins at conception", she says.

Widcombe's position is a simple one to understand. The problem faced by the pro-research lobbyists is explaining to MPs the more complicated concept that the beginning of the embryonic period proper is when the first cells which may form the embryo itself, rather than its supporting structures, can be identified, at 14–16 days from conception. Pro-research groups reject the use of the term 'pre-embryo', by PROGRESS and the MRC, to describe the embryo in the 14 days following conception. They say the term is deliberately misleading and Braine's pamphlet charges that the term 'embryo' has been shunned during lobbying, despite its use in a professional context by researchers. The MRC says it is standing by its use of 'pre-embryo', although Mary Rice, who is co-ordinating the MRC's campaign, acknowledges that the term has caused some confusion.

Thurnham believes that the most effective tactics for pro-research lobbyists may be simply to explain to MPs the medical benefits that are promised by embryo research.

If research is banned, some scientists predict an exodus of British researchers abroad, a fear first voiced by Sir George Porter, president of the Royal Society. But Jenny Gunning, who has compiled a report comparing embryo research internationally, disagrees. She predicts that most British researchers would not emigrate, but would choose to collaborate in projects where experiments can be carried out in other countries.

Winston also thinks a ban on research would not lead to the loss of many researchers from the United Kingdom. Embryo research is "a small, but important" part of his own group's work, and most researchers could occupy themselves in other areas, he says. Winston adds that the pro-research arguments should concentrate on future benefits to patients, not the effects of a ban on research on scientists.

Internationally, there is no consensus on the embryo research question. Belgium and the United States have no legal restrictions, although the US government gives no financial support for the work. Sweden allows research up to 14 days from conception, setting a precedent for the British pro-research lobbyists, but a bill banning research in West Germany is expected to become law in the summer.

Peter Aldhous

Genentech wins round two

Washington

GENENTECH, the biotechnology company based in South San Francisco, gained a victory in its long-running battles with the UK-based Wellcome Foundation and Genetics Institute of Massachusetts when a jury in a district court in Delaware ruled last week that three of its key patents for the blood-clot dissolving agent tissue plasminogen activator (TPA) had been infringed.

The court case is almost a re-run of the action brought by Genentech in 1987 to prevent Wellcome from infringing on its TPA patents in the United Kingdom (see *Nature* 328, 189; 1987). The outcome was, however, quite different, as the British High Court declared Genentech's wide-ranging TPA patents to be invalid. Genentech appealed, but the decision was upheld in the British Appeal Court. The outcome of these quite separate court actions outlines the dissimilarity between

AWARDS

Australia's new prize

Sydney

THE Australia Prize, a new international science prize worth A\$250,000, has been awarded to Eugene Nester of the University of Washington in Seattle, Jeff Schell of the Max Planck Institute and Allen Kerr of the White Agricultural Research Institute in Adelaide for their work on crown gall disease.

The prize was created as part of an initiative by Barry Jones, until recently the Minister of Science, to raise the profile of science in Australia. It will be awarded in a different area of science each year and only those who have not already received a leading award in the nominated field will be eligible to receive it.

The 1990 Australia Prize was awarded for excellence in the field of biological sciences related to agriculture and the environment. Crown gall disease is responsible for annual losses of A\$100 million in damage to crops worldwide.

Kerr isolated the bacterium, *Agrobacterium tumefaciens*, responsible for crown gall disease. Nester and Schell have studied the way in which the organism causes the disease and have developed a genetically engineered benign strain of the bacterium.

Although the prize was established to boost the prestige of science, it has not been popular with the Australian scientific community. The administrative costs of the prize came to A\$160,000 in its first year of operation and despite assurances from the government that this money would not come out of the science budget, many scientists feel that support for research would have been preferable to support for a prize.

Tania Ewing

US and UK biotechnology patent law. It is like comparing "apples and oranges", says Susan Rogers of Genentech.

The Delaware action is seen as an attempt by Genentech to stave off the competition from Duteplase, Genetics Institute's version of TPA that is licensed to Wellcome, as well as other 'second-generation' TPAs now under development. Wellcome is currently awaiting approval of Duteplase by the US Food and Drug Administration.

Although the judge has not entered a final ruling on the jury's decision, Genentech's president and chief executive officer, G. Kirk Raab was quick to claim that "this verdict is a major victory for the biotechnology industry because it signals that such patents will provide meaningful protection for biotechnology innovations".

The action against Wellcome and Genetics Institute was brought by Genentech and was countered by Wellcome's anti-trust action and Genetics Institute's claims of unfair competition. The jury ruled that both these claims were without merit. Although the jury upheld the validity of the three patents, no monetary damages were awarded to Genentech. Wellcome spokesman Martin Sherwood would not comment on the jury's decision until the judge has entered a final ruling.

Genentech's TPA, marketed under the trade name Activase, is the only form of TPA approved for sale in the United States for the treatment of heart attacks. In 1989, sales of TPA reached \$196.6 million, an increase of 30 per cent on the previous year. It is estimated that the worldwide market for thrombolytic agents will reach \$500 million by 1993. Susan Rogers, a spokeswoman for Genentech, says that Genentech will be filing an injunction against the two companies to prevent them from marketing Duteplase in the United States. At this stage, it is unclear how this decision will affect the 'second-generation' TPAs such as novel plasminogen activator, which Genetics Institute licensed to Burroughs Wellcome, the US subsidiary of Wellcome, in December 1989. The drug is in the pre-clinical phase of development and is a modified form of TPA with deleted fibronectin and epidermal growth factor domains and changes to the glycosylation sites. It is hoped that this modified form of TPA will have a longer half-life in the bloodstream than TPA so that it can be given as a bolus injection over a shorter period of time.

The Delaware court's decision will provide a welcome boost to Genentech, which is fighting to hold on to two-thirds of the US market share. Last month, results of an Italian direct clinical trial between TPA and streptokinase (Hoechst-

Hopes of better times

São Paulo

THE Brazilian Congress has turned into law a provisional measure reducing restrictions on the import of scientific equipment and laboratory materials. The measure was among the last issued by the government of ex-President José Sarney and was strongly supported by scientists.

The new law was unanimously approved by all parties. There was a last-minute threat when one senator wanted to bar the import of equipment when a Brazilian 'alternative' existed. But lobbying defeated the amendment on the grounds that it would have continued the practice of forcing scientists to buy Brazilian regardless of quality.

The new secretary of Science and Technology, physicist José Goldemberg, also announced recently that he is negotiating loans from the World Bank and from the Inter-American Development Bank to help science and technology projects. Goldemberg is thought to be asking for \$400 million.

Nobel prizewinning physicist Abdus Salam is also trying to encourage Brazil to spend more on science. Salam, who is head of the International Centre for Theoretical Physics at Trieste and has been active in encouraging international banks to help developing countries to invest in science, last month wrote to Brazil's new president, Fernando Collor de Mello, saying that Brazil's investment in science is too low given the size of its economy. Brazil spends roughly 0.7 per cent of its gross domestic product on science and technology.

Goldemberger wants to try to double that figure by the end of Collor's five-year term. But the previous president's science ministers set a goal of 2 per cent — and failed.

Ricardo Bonalume

Roussel Pharmaceuticals) showed that TPA is no more effective than streptokinase, which is one-tenth of the price and the preferred drug in Europe (see *Nature* 344, 183; 15 March 1990). In addition, Genentech is having to fend off new competition from SmithKline Beecham's Eminase, which has a longer half-life than TPA and can therefore be administered as a bolus injection over 2–5 minutes, instead of being infused over a 3-hour period as in the case of TPA.

Stewart Adkins, stock analyst for Shearson Lehman Hutton in the United Kingdom, believes that Duteplase is not going to be particularly important for Wellcome, which is better known for the AIDS drug, AZT. In view of the fact that aspirin and streptokinase is a cheap combination and that the mortality data are favourable, he feels that there is not a lot of money to be made in the thrombolytic market and the likelihood of making a return on investment is diminishing.

Diane Gershon

SPACE

Japan plans new lunar mission

Tokyo

SCIENTISTS at Japan's Institute of Space and Astronautical Science (ISAS) are planning an ambitious follow-up to their recent success in sending a spacecraft to the Moon (see *Nature* 343, 403; 1 February 1990). Last week, a committee at the institute decided to give priority to a mission to monitor 'moonquakes' by dropping penetrators on the Moon from an orbiting spacecraft.

The institute's next lunar mission is tentatively scheduled for 1995 and will use the new larger M-V solid-fuel rocket currently under development by the institute. It will be capable of carrying much larger payloads than its present launch vehicle, the M-3SII.

The spacecraft will drop three 13-kg, 12-cm diameter penetrators, each about the length of a cricket bat, from orbit onto widely spaced locations on the lunar surface: one near a former Apollo landing site, one on the far side of the Moon, and another near the northern lunar pole. Retro-motors fitted in the penetrators will reduce the speed of impact, but it will still be a 'hard landing' at a speed of 250–300 metres per second, according to Hitoshi Mizutani of ISAS.

Mizutani says the penetrators will plunge 1–3 metres into the lunar regolith (topsoil). Tiny seismometers and other instruments carried in the penetrators will then relay information on 'moonquakes' and heat flow to Earth via the orbiting mother spacecraft. The instruments are

expected to keep working for about one year.

ISAS scientists have been perfecting the design of the penetrators by firing life-sized models into artificial lunar soil with a high-speed gun at their Noshiro Rocket Center in northern Japan. Details of the technology developed for protecting the sensitive seismometers from lunar impact are "secret", says Mizutani. But basically the instruments will be packed in light epoxy material and great care will be taken to ensure that there are no moving parts that could be displaced as the penetrators hit the Moon at close to the speed of sound.

The plan has to cross several hurdles before it wins approval and funding from the Ministry of Education, Science and Culture (MESC), to which ISAS is affiliated. According to Jun Nishimura, director general of ISAS, the project will be screened by another couple of committees before official ISAS approval can be granted in June. It will then be for MESC to decide if funding will be requested from the Ministry of Finance.

The lunar project was one of three being considered by ISAS for launch in the mid-1990s. The others are a mission to drop a balloon into the atmosphere of Venus and an ambitious proposal to collect dust from the comet Wirtanen and return it to Earth. The lunar mission is being given priority and the others will be taken up for discussion next year.

David Swinbanks

HUBBLE TELESCOPE

Astronomers' wait continues

Greenbelt, Maryland

A STICKING valve in the Space Shuttle's hydraulic system stopped the long-delayed launch of the Hubble space telescope four minutes before its scheduled take off last week. Replacing a key part is expected to take one to two weeks, with the next launch attempt scheduled for 25 April, the National Aeronautic and Space Administration (NASA) said.

The shuttle and its \$2,000 million payload will remain on the launch pad at Kennedy Space Center during the repairs. NASA's main worry now is that the telescope could be contaminated as technicians work on the shuttle and repair the telescope's batteries. Air pressure within the shuttle's cargo bay is maintained at higher than atmospheric level to keep dust and other contaminants out.

But to recharge the batteries, technicians will need to enter the compartment, which is normally maintained at near clean-room standards to prevent the entry of dirt or even insects that could

show up as specks in the telescope's sensitive optics. The extra two weeks on the ground is nevertheless expected to increase contamination of the Hubble's mirror by about 0.1 per cent.

NASA officials say they will surround the cargo bay with a form-fitting shell during the repairs. Although air pressure within the bay will drop as the bay doors are opened, NASA expects that the airtight gantry will keep pressure in the cargo area above atmospheric levels, and thus relatively free of contaminants.

The faulty valve is part of the first auxiliary power unit, one of three hydrazine-powered engines that run the turbines that maintain the shuttle's hydraulic pressure. When NASA technicians install a replacement the power unit this week, it will be the first time that such a procedure has been attempted on the launch pad, with the shuttle in a vertical position. Because the operation is unfamiliar, NASA is allowing extra time for its completion.

G. Christopher Anderson

GRAVITY WAVES

SERC's saving is Scotland's loss

London

AN Anglo-German review panel has given preliminary approval to a plan to build a gravity-wave observatory, which could be completed as early as the mid-1990s. But to the disappointment of the British teams involved in the project, the observatory seems likely to be built in West Germany, rather than in Britain. The project is expected to begin after formal approval is granted by the UK Science and Engineering Research Council (SERC) in July.

British physicists hoped that the observatory would be built in Tentsmuir forest, Fife, in Scotland. Although the site is excellent, West Germany will have the right to choose the observatory's location as SERC plans to contribute only about £5 million towards the project's estimated £30 million costs. German funding is expected to come from the Research Ministry (BMFT), the local *Land* and the Max Planck Society.

Professor Jim Hough, from the University of Glasgow, says that his group (which designed the observatory together with a German team from the Max Planck Institute for Quantum Optics at Garching) had hoped that SERC would pay for a large part of the project, but were told informally about a year ago that most of the money would have to come from abroad.

Hough still hopes the detector can be built in Scotland, should any problems arise in Germany. The first German site to be selected, in Bavaria, was abandoned because of difficulties in gaining planning permission. Although approval for a site just north of Hanover should be easier, German physicists are still waiting for the final word from the local authority. Hough's team has obtained planning permission for the Tentsmuir site.

The observatory will consist of four large suspended masses, at the end of two 3-km-long arms, arranged in an 'L' shape. The interference patterns of laser light bouncing between mirrors on the masses will show up extremely small changes in their spacing — changes that could be caused by ripples in space-time as predicted by the theory of general relativity.

These ripples, or 'gravity waves', should arise from the movement of enormous masses — for example the collapse of a star's core in a supernova explosion. Observing these gravitational waves has been described as "one of the last-known windows that remains to be opened on the Universe" (see *Nature* 321, 378; 1986).

The US National Science Foundation also intends to build gravity-wave detectors and has plans for a pair of 4-km arm facilities to be located at still-undecided sites in the eastern and western United States. If Congress agrees to fund the project this year then the detectors should be in operation by 1995.

Peter Aldhous

Dollars are green too

Washington

US administration officials promised last week that they would unveil "concrete proposals", including an international global change research programme, at the first White House international conference on global climate change, which is being held in Washington this week. The meeting, to be opened with a major environmental policy address by President Bush, has been presented as an opportunity to address the economic issues that the administration believes have been overlooked in global climate change negotiations.

But foreign participants complained that the agenda of the meeting seemed oriented towards further research instead of action, emphasized the economic costs of countermeasures and repeated many of the issues already being discussed within the United Nations, which is the designated central organizing body for global climate change policy making. And in a week that ends with Earth Day, environmentalists converged on the meeting to lobby for tough cuts in greenhouse gas emissions, rather than "stalling" with more research.

White House science adviser D. Allen Bromley last week outlined a proposal for collaborative research initiative "comparable on the international scene to the US national programme". The collaboration could include a joint-funding pool for global change research, coordination of existing research programmes, and backing for Third World projects, such as Indonesia's proposal to build a giant radar to explore the little-understood equatorial atmosphere.

The conference, which fulfils a promise from George Bush's 1988 presidential campaign, was first announced at the Malta summit last December. Since then, US planners have rushed to try to assemble an agenda that will not duplicate the policy-making process of the UN Intergovernmental Panel on Climate Change (IPCC).

But as ministers from 17 nations, the European Community, and the Organization for Economic Cooperation and Development gathered in Washington, diplomats from the invited countries said they were puzzled by the preliminary agenda, which promised discussion on issues such as "how well can we predict temperature trends?" and "how 'good' are our global [climate] models?" Questions of that sort are already the purview of the IPCC, where a panel chaired by Britain is preparing to release its own scientific analysis in May (*Nature* 344, 577; 12 April 1990). Environmentalists suggest that the meeting is an attempt to get a "second opinion" in anticipation of ominous con-

clusions from the IPCC. "In essence, the focus of the White House conference is akin to saying 'we don't like those projections'", Alden Meyer of the Union of Concerned Scientists told Congress earlier this month.

White House officials hope that the conference will spur new economic research to determine the real costs of global warming response measures such as energy-efficiency standards and "carbon taxes". "We feel that the economic aspects have not been studied to the degree that they might", says Michael Deland, chairman of the White House Council on Environmental Quality.

While many economists believe that the strict controls on emissions of carbon dioxide now advocated by many European countries will cost more than they will save in energy efficiency, environmentalists claim that going green will save money. But hard numbers are in short supply.

Bush administration officials have been outspoken in their belief that a 20 per cent reduction of US carbon output by early in the next century would be extraordinarily expensive — costing as much as \$3.6 billion million according to one estimate — but they admit that there have been almost no reliable economic studies to back that up.

In part to ascertain the current state of knowledge on the economics of global warming, the White House asked the 18 visiting delegations to fill out a 23-question survey on the scope, details, and conclusions of their economic and scientific research on climate change. But several countries have reacted badly to the questionnaire, saying that the questions appeared to be slanted towards the high economic cost of dealing with global warming.

At least one nation — West Germany — refused to answer the questionnaire, objecting to "leading questions" that implied a choice between research and response measures. A West German position paper submitted in the survey's place stresses the need for immediate action.

Besides questions about the amount, types and mechanisms of global warming research in the invited countries, the US survey asked the delegates to identify "possible conflicts, if any, between your interest in continued economic progress and your interest in arresting possible undesirable global change".

West German chancellor Helmut Kohl wrote to Bush earlier this month asking that the White House modify the conference's agenda to place greater emphasis on practical policy responses, rather than further research. But US officials say that policy moves at this point could preempt

US eases controversial visa restrictions

Washington

IN an attempt to quell mounting opposition to its restrictions on entry into the United States for people with AIDS, the Bush administration last week announced a new no-questions-asked, 10-day visa for visitors attending certain US scientific and medical conferences.

Rather than stating that they are infected with the AIDS virus and answering a series of questions to determine the risk they present to the public, people with AIDS who are going to designated US conferences will be able to obtain a visa simply by giving the name of the conference and satisfying normal entry standards, the Immigration and Naturalization Service said. The first two conferences for which the special visas will apply are the Sixth International Conference on AIDS to be held in San Francisco in June and the Congress of the World Federation of Hemophilia, scheduled for Washington in August.

But many AIDS activists complain that the measure does not go far enough. They believe that AIDS should be removed entirely from the list of contagious diseases, as the US Centers for Disease Control recently recommended. AIDS had been placed on the list in 1987 by conservative Senator Jessie Helms (Republican, North Carolina) with the effect that carriers had limited rights of entry into the United States. As of last week, about one-quarter of the 12,000 attendees expected at the San Francisco conference had decided to boycott the meeting because of the visa restrictions. AIDS activists said they did not expect the new measures to change the boycott picture substantially.

G. Christopher Anderson

ongoing studies at IPCC. Although the Bush administration "does not believe that further research is any substitute for action", Bromley said, it is committed to the IPCC aim for an international "Framework Convention" to set global emission levels, rather than starting an international policy-making process of its own.

US officials say that this week's conference will be different from the meetings of IPCC in that it will bring together the scientific and economics aspects of global warming.

Among the attending nations are the Soviet Union, Japan, Brazil, the United Kingdom, Australia, Canada, Poland, and Indonesia. Although Bush had previously promised that China — the largest producer of greenhouse gases — would be invited, diplomatic relations between the two countries have soured considerably since the 1988 presidential campaign.

G. Christopher Anderson

ARIANE

Dirty rag costs Arianespace forty million dollars

Paris

A DISCARDED rag caused the explosion which wrecked flight 36 of Europe's Ariane-4 commercial launcher in February (see *Nature* 344, 7; 1 March 1990). Tests on fragments of flexible tubing salvaged from the wreck are still being carried out by experts at the Saclay Jet Propulsion Test Centre near Paris, but the discovery that a main water valve in the cooling circuitry had become blocked by "textile material" confirms one of four principal conclusions of a 180-page report made public last week.

The report follows investigations by a seven-person specialist board of enquiry set up by the European Space Agency (ESA) and Arianespace, the Paris-based launch consortium, immediately after the incident. According to Arianespace chairman, Frédéric d'Allest, Ariane launches should now be able to resume this summer.

On the basis of initial telemetry data, the loss of the mission was attributed to a decrease in thrust in engine D, one of the four first-stage Viking motors. This drop occurred 6.2 seconds after ignition and, d'Allest told the press last week, was the result of "an almost total obstruction of the water feeding circuit of Viking motor D". The obstruction occurred upstream of the motor, before the water pump, and the motor itself is not under suspicion.

The board last week felt the probable cause of the obstruction was either the "untimely presence of a foreign object in the water pipes or a failure in the main water valve". The water pipes are part of the cooling system of the turbopumps in the Ariane first stage. But the latest tests at Saclay have now shown that the valve itself was not to blame.

The discovery of the rag follows a remarkable and hazardous recovery operation. Working for almost 1,600 hours in total darkness in mangrove swamps and deep water, teams of divers recovered over 450 pieces of the disintegrated rocket. The divers were aided by legionnaires, Paris fire brigade officers from the launch centre and scientists of the Grenoble Research Centre (LETT) equipped with highly sensitive detection apparatus. On 9 April most of the water tank and 4 to 5 metres of water feeding pipe were recovered and were sent to France the next day for analyses.

The question now taxing Arianespace is how the rag went unnoticed. A company press communique says "the most likely cause for the presence of this textile material is tied to unusual servicing on this tubing". On Monday, an Arianespace

spokesman said that the tube had initially been installed on another rocket, but was not needed. It was taken off and stored until finally being reinstalled on flight 36. It was this "exceptional" procedure, he said, which had allowed the rag to go unnoticed.

This points to a weakness in safety checks following very unusual procedures, he said.

The investigation has so far also revealed that there was an "incipient fire" in the propulsion bay of one of the liquid strap-on boosters (PAL 3) 2.4 seconds after firing of the motors.

This was caused by leaking nitrogen peroxide propellant but, says the board of inquiry, the two incidents do not seem to have been related.

The board, chaired by Jacques Durand, head of the Arianespace programme at ESA, has put forward 44 recommendations, nine of which, they say, must be acted upon before flight 37 receives authorization.

The recommendations mainly concern tightening up verification procedures during launch integration and additional checks on liquid coolant circuits. Extra thermal blanketing is likely to be used on sensitive circuits exposed to a fire risk, so that a small fire would not jeopardize the entire mission.

The loss of flight 36 has so far cost Arianespace between FF200 and FF300 million (\$36 to \$54 million), but d'Allest says that clients still have confidence in Ariane's reliability. Four new launch contracts have been signed since the crash: with the US company Hughes Communications, the Berlin Institute of Technology and France Telecom. D'Allest thinks launches should resume in the summer and that four launches could still be carried out this year.

Priority in the future will be given to the Japanese customers, the Japan Broadcasting Company (NHK) and Space Communication Corporation (SCC) who lost their telecommunications satellites in the crash. Both satellites were built by US manufacturers. SSC's Superbird-B satellite was made by Ford Aerospace for television and defence use.

The other satellite, BS-2X was built by General Motors at a cost of \$93 million, to boost NHK's satellite television broadcasting capability. Now the company still has only one satellite in operation, Yuri 2B, which is expected to stop functioning in about a year. The future of the satellite service now depends on the successful launch of the follow-up to Yuri-2B, BS-3, by Japan's National Space Development Agency, in August.

Peter Coles

INDUSTRIAL RESEARCH

CBI goes critical

GOVERNMENT support for British industrial research and development needs overhauling, according to a report* published last week by the Confederation of British Industry (CBI). Through the Department of Trade and Industry (DTI)'s Enterprise Initiative, about £100 million of public money is spent annually on industrial research and technology transfer, in fields from aerospace to biotechnology.

The CBI report argues that the process of applying for DTI research grants needs to be streamlined, since "some companies claim that additional management costs incurred as a result of grant application now equal . . . the grant itself".

Small companies also need more advice on how to implement existing technology, the CBI says. The report suggests that a separate "technology consultancy" could be set up within the DTI to promote technology transfer.

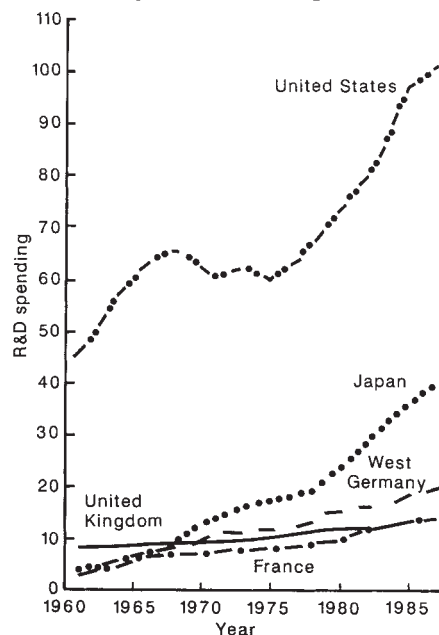
P.A.

*Technology and Enterprise — a CBI review of DTI support for technology in UK industry.

RESEARCH SUPPORT

Japan catching up

FIGURES from the US National Science Foundation (NSF)'s National Science Board biennial report Science and Engineering Indicators — 1989 show that the United States still spends more on research and development than the combined efforts of the next four nations — Japan, West Germany, France and the United Kingdom. Over the past decade,



there has been virtually no change in the US position of dominance — now 54 per cent of the total expenditure of the five nations. But the huge gap between the United States and its principal high-technology competitor, Japan, narrows greatly when only non-military research is considered.

A.A.

Back-door trial for US?

Washington

THE controversy over the French abortion pill RU-486 looks set to reopen in the United States following an announcement by California state officials that they are making plans for the first US tests of the drug. But California's efforts may prove to be a vain political gesture — the manufacturers of RU-486, Roussel Uclaf of Paris, are unwilling to ignore continuing opposition from US anti-abortion groups and supply the pill for clinical trials.

California's Attorney General John Van de Kamp is urging the State Health Department's Food and Drug Branch (FDB) to use a 1987 state law introduced for speeding tests of AIDS drugs to initiate rapid clinical trials of RU-486 in California. The state law would enable Roussel Uclaf to by-pass the long process of approval by the Food and Drug Administration (FDA) through which most US drug licences are granted.

Although RU-486 has yet to be tested in the United States, it has been used by more than 40,000 women in France, and is undergoing trials in more than a dozen other countries. Roussel Uclaf plans to market it in the United Kingdom, the Netherlands and Scandinavia.

NASA EXPERIMENT

Return of the mutant tomatoes

Washington

HUNDREDS of US schoolchildren who had planned to grow tomato plants from seeds that had spent six years in orbit switched to the terrestrial variety last week after newspapers quoted a government memorandum that suggested the space-borne seeds could be dangerous. The NASA memorandum stated: "There is a remote possibility that radiation-caused mutations could cause the plants to produce toxic fruit".

NASA's Long Duration Exposure Facility carried 12.5 million of the seeds until its return in January. Packets of the seeds were sent to 58,000 teachers nationwide as part of a programme to encourage children to become interested in science and space by participating in a simple experiment comparing the growth of normal tomato plants with those exposed to six years of cosmic radiation.

NASA had not anticipated that the children would be tempted to eat the fruit, and the included instructions carried no warning against consuming the experiment. At a press conference late last week NASA said that the seeds were exposed to less radiation than many earth-grown varieties and would be "as safe as any fruit you buy at your local grocers".

G. Christopher Anderson

The State Health Department's director Ken Kizer says that under the current law California would authorize testing of RU-486 if it received an appropriate application from its manufacturers, but that the speed of a clinical trial would depend on the availability of State financing.

John Van de Kamp, who wrote the 1987 legislation, has already promised financial support for clinical trials of RU-486 if he is elected Governor of California in the January 1991 election. Researchers at the University of California Medical School, led by Marcus Conant, are now preparing a trial protocol which they hope to submit for State approval.

The Attorney General's office is optimistic that their moves will persuade Roussel Uclaf to apply for clinical trials. "The consensus is that that [Roussel Uclaf] will not be looked on favourably at the FDA and what they want is a favourable regulatory climate", said an office spokesman. "We would welcome a trial application, process it as expeditiously as possible and get the tests under way", he added.

The Planned Parenthood Federation of America is more sceptical. Louise Tyrer, its vice-president for medical affairs, said that they along with a "parade" of other pro-choice groups had met with little success in persuading Roussel Uclaf to market RU-486 in the United States, and that the moves for independent Californian trials stem more from political posturing than "reality".

At present, the United States could obtain RU-486 only through the World Health Organization (WHO) which funded some of Roussel Uclaf's research and has a contract to sell the drug at cost price. But WHO's first concern is with the developing world and is an unlikely source of RU-486 for California.

Following events in France, where Roussel Uclaf at one stage withdrew RU-486 in the wake of threats from the US Right to Life Committee to boycott Roussel-Uclaf products and those of Hoechst, the majority share-holders (see *Nature* 340, 6: 6 July 1989), the company's directors are nervous about supplying the drug for tests in the United States.

Madame Mouttet, responsible for marketing RU-486 in France, said that claims that clinical trials are to begin in California are "not true". "To carry out trials, they would need supplies of RU-486 and Roussel Uclaf has no intention of providing the drug". She says the company has no intention of marketing the drug in the United States for the foreseeable future — at least until the "political and social" problems associated with abortion in the United States are ironed out. **David Concar**

New suit targets university laboratories

Boston

JEREMY Rifkin, head of Foundation on Economic Trends in Washington, DC, is attempting to continue his run of winning lawsuits with new action designed to stop projects paid for by the Pentagon Biological Defense Research programme at some fifty universities and research institutes in the United States.

Rifkin has been fighting biological weapons research for half a decade. In 1985, a suit brought by Rifkin's organization halted the construction of a controversial aerosol testing facility in Dugway, Utah. In 1987, Rifkin forced the US Defense Department to conduct an environmental impact assessment of the entire Biological Defense Research Program. With that assessment complete at the end of last year, Rifkin is now back in court, calling the document "grossly inadequate" and trying to force the programme to close as a health and environmental threat.

The new legal case, filed on 4 April in federal district court, calls for projects rated as of higher risk by the Pentagon to be suspended until individual environmental assessments are conducted for each of them. Charles Dasey, an Army spokesman, says the Army will not comment on the pending lawsuit, but stands by its decision last December to continue the \$60-million programme. Dasey noted that the Army's official decision, after completing its environmental assessment, was to continue its funding because of the "overriding national importance of the programme and the insignificant effect of the existing programme on the quality of the human environment". Yale Medical School's Arbovirus Research Unit comes in for particular criticism from Rifkin. The laboratory, which receives funding from the Pentagon's Biological Defense Research Program, conducts research on arboviruses which are borne by insect-vectors and cause diseases such as yellow fever. Rifkin, who believes that research on such deadly diseases needs to be better protected against the threat of sabotage or accidental release of the infectious agents, says the Yale laboratory is "probably the most dangerous single university lab in the country," and alleges that the laboratory safety procedures and emergency protocols are "almost a joke." Gregory Tignor, who heads the Arbovirus Research Unit, has said in the Yale university newspaper that he takes "personal offense" at Rifkin's accusations. Tignor, who notes that his unit has never had an accident, says the laboratory is "serious" about safety issues. But Tignor also says that he would probably apply only for National Institutes of Health grants in the future. "I don't want the controversy," he said.

Seth Shulman

AUSTRALIAN ARCHAEOLOGY

Emphasis on 'aborigine rights'

Sydney

FURTHER protests in Britain are being planned by the Tasmanian Aboriginal Centre following their success last month in obtaining the return of the mummified head of a Tasmanian Aboriginal from the Royal College of Surgeons in Dublin. Bob Weatherall and Michael Mansell, two representatives from the centre, have already mounted one protest in February outside the Natural History Museum which holds a large collection of aboriginal remains.

The head held in Dublin was that of the grandfather of Weatherall and was returned to him by the Australian Embassy in Dublin. The college had actually intended that the head go to the National Museum of Australia in Canberra and not to the Tasmanian Aboriginal Centre, according to Joseph Grace, Chief Administrator for the college. But the Australian Embassy in Dublin says that the Australian government supports the return of remains to their ancestors.

"Custodianship of remains is vested in the National Museum of Australia until such time as the appropriate aboriginal community is identified and accepts them. In this case Mr Weatherall was clearly related and there was no need for the remains to undergo an interim period at the museum", an embassy spokesperson said.

Although citizenship was only granted to Aborigines in 1967, the federal Aboriginal and Torres Strait Islanders Heritage Protection Act of 1984 provides for the protection of sacred objects, skeletal material and sites of significance. The federal act overrides legislation by the state and territories.

According to Fred Behr, spokesman for the Department of Aboriginal Affairs, the federal act can be used when the aboriginal community is unhappy with state legislation. "The federal minister can, in consultation with the aboriginal community, place a protection order on a site which overrules state legislation." The National Museum of Australia plays a safekeeping role until the federal minister, together with the aboriginal community, decides what to do with the remains.

That places the National Museum of Australia in a very different position from the British Natural History Museum which is prevented from giving up its collections by the British Museums Act of 1963. A museum spokesman said at the time of the protest in February that it "would be a tragedy" to devalue the collection by dispersal.

In Australia, some collections are already being returned. In 1988, five aboriginal communities based around the Victoria-New South Wales border requested the return of the George Murray

Black collection from the Victorian Museum and the National Museum of Australia. The collection, the largest in Australia, was assembled in the 1920s and 1930s; part of it has already been returned and a spokesperson for the National Museum of Australia said the remainder will be returned "in the near future".

Australian policy has been hard on skeletal archaeologists who have abandoned the field in increasing numbers over the past decade. Peter Brown, a lecturer in palaeoanthropology at the University of New England in Armidale, is one researcher who left the study of Australian bones after finding it increasingly difficult to do fieldwork in Australia. Brown now works exclusively on remains in China and Europe and is encouraging his students to do the same. "Our (Australian) museum collections are poorly documented and it is impossible to do field work in this country. . . . Australia has a limited future as far as skeletal archaeology goes and there's no point training students in it if they can't get jobs." But Colin Pardoe, who is one of only five

skeletal archaeologists left in Australia, has a different view. He is a research fellow at the Institute for Aboriginal and Torres Strait Islanders Studies and says that any animosity between aboriginal communities and archaeologists arises out of the failure of archaeologists to return information, gained from the investigation of remains, to the aboriginal relatives.

Pardoe agrees with Brown that there is a feeling of pessimism in Australian archaeology. "Students are generally advised not to go into skeletal studies. There is a perception that soon there won't be anything to study. However, the future of Australian archaeology is clear. It involves the reality of aboriginal ownership, the continuation of cooperation and collaboration and the sharing of information in a way that can be understood by the aboriginal community", Pardoe said.

According to Jim Berg, coordinator of the Koori Heritage Trust in Melbourne, it is only white students who are being discouraged from pursuing skeletal studies: "We have our own people coming through the system (studying anthropology). We'd like our own history to be told by our own experts."

Tania Ewing

MARINE BIOLOGY

Stormy move for laboratory

Washington

'AQUARIUS', the world's most advanced underwater research station, is to be moved from the sea floor off St Croix in the US Virgin Islands because of damage inflicted by Hurricane Hugo.

According to the National Oceanic and Atmospheric Administration (NOAA), the storm destroyed the laboratory's surface support buoy and some onshore research facilities as it devastated the Caribbean last September and has left St Croix without critical support services, such as emergency medical care. But scientists involved with Aquarius say that the decision to move it was politically motivated and could hurt marine science in the region.

Aquarius is an 81-ton, trailer-like structure with a laboratory and living quarters that allows teams of up to six aquanaut-scientists to live and work on the sea floor for as long as two weeks. Typically stationed at a depth of 50 feet, the laboratory allows scientists to make scuba excursions to depths of 130 feet for as long as six hours at a time. Ordinarily, divers would be forced to restrict their dives to a few minutes or to undergo long decompression periods in order to return to the surface after dives to that depth.

Aquarius was first deployed in 1987 with the intention that it would eventually be moved. To make the project more attractive to the rest of Congress, the pro-

ject's sponsor, former Senator Lowell Weicker (then a Republican from Connecticut), arranged for Aquarius to be movable to other marine sites. But now that the project is actually to relocate for the first time, marine scientists are concerned that valuable science opportunities will be lost. "If it were a strictly scientific decision, it would stay where it is", says Elizabeth Gladfelter, director of the West Indies Laboratory, which oversees Aquarius, "the laboratory was just realizing its full scientific potential at the site."

NOAA officials are considering other sites in the Caribbean, the Bahamas or the Florida Keys. The move is expected to take between six months and a year once a final destination site has been selected. But none of the sites contemplated is likely to provide the scientific opportunities of the current Salt River location, says former project director John Ogden. The existing site is on the floor of a submarine canyon with extensive coral reefs on both walls. Biologically interesting sites that are both deep and close to shore are relatively rare, he says.

Access to the West Indies Laboratory — regarded as the best-equipped marine biology lab in the Western Hemisphere — is also seen as a significant loss. Despite \$1.5 million in damages during Hurricane Hugo, the laboratory is now back in operation while repairs continue.

G. Christopher Anderson

When is a drug 'safe'?

SIR—The Food and Drug Administration (FDA) is often criticized (as in a recent leading article in *Nature* 343, 494; 1990) for its handling of various issues or specific products. While some criticism implies that the agency is too lenient on regulatory policies or decisions, other criticism takes us to task for being too slow. A favoured theme, one implicitly endorsed in the leading article, is "once any amount of possible effectiveness has been demonstrated, the FDA should approve the product immediately, making it available to patients who need it" (although *Nature* expressed more concern about the health of the manufacturers than about that of the patients). These critics are right to be critical of any genuine delay in bringing a safe and effective drug to market, and the FDA shares their eagerness for a streamlined drug-approval process: there is, indeed, a genuine cost of regulatory delays. However, any patient—even one with a fatal disease—can be made acutely worse. Thus, the costs of time in the review process must be weighed against the hazards of approval of an unsafe, ineffective drug.

The FDA's regulatory approach will do far more to bring safe and effective new drugs sooner to patients who need them than if it were to grant marketing approvals based on initial results. In fact, this latter path would have it ignore both the laws that ensure the safety of new drugs and the realities of clinical research.

Clinical researchers are aware that the earliest reports on a new drug are often over-positive. That is, initial reports can be more optimistic about how safe the drug is and how well it performs its intended purpose than later studies prove it to be. Consequently, the FDA and the medical community usually withhold judgement until some verification is obtained.

The FDA has already used its existing regulatory procedures to evaluate and approve numerous products of the new biotechnology, with many more in the pipeline; some 400 (the majority of which are diagnostic test kits) are approved for marketing, and more than 700 drugs and biologicals are in clinical trials. Marketing approval times for the new biotechnology-derived drugs and biologicals have averaged approximately half the 33 months required for new products generally; we do not, indeed, mindlessly crank each and every product "through the same rigorous and probably over-rigorous mill". The agency is committed to refining further its procedures for new drug evaluation. Recent changes in FDA regulations should make investigational drugs available to patients who require them at an earlier time than previously

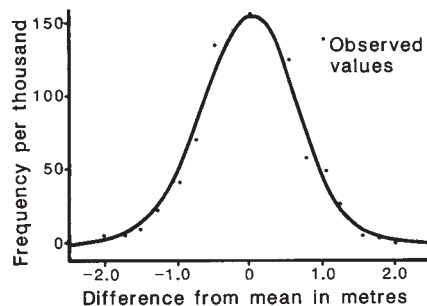
and should shorten review times. The FDA's decisions exert a profound impact on both patient care and commercial development, and it is imperative that the agency's actions be based on—and only on—solid scientific analysis. A high-quality, predictable review process fuelled by a strong research base can encourage the development of promising new therapies and speed the appearance of new, safe products. The drug review process is a complex and difficult one, and approval of an unsafe drug, or unnecessary delay of a good drug, can each have disastrous consequences. The FDA must avoid both.

HENRY I. MILLER

Office of Biotechnology,
US Food and Drug Administration,
Rockville, Maryland 20857, USA

Identifying trends

SIR—In his leading article entitled "Two gales do not make a greenhouse", John Maddox (*Nature* 343, 407; 1990) made the cautionary comment that "even indirect and properly qualified assertions that transitory events are evidence for more durable physical phenomena may back-



Maximum annual gauge reading at Roda Gage, Cairo, Egypt for 1,080 years between AD 641 and AD 1946.

fire". H. E. Hurst¹ made just such a point as scientific consultant to the Ministry of Public Works, Egypt, in 1950 when he studied the discharge of the Nile river at Aswan. It appears that the Egyptians were particularly diligent about keeping records of the annual flood levels of the Nile at Cairo. The distribution of over one thousand readings (see figure) is well fitted by a normal curve, suggesting that the magnitude of the flood level was a random event. However, this long series of records of flood levels at Cairo showed a tendency² for groups of high or low values to occur. Hurst remarked that such a tendency is greater for natural than for truly random events.

I have no knowledge of the reaction of the Egyptians to a series of high annual flood levels at Cairo but it is possible that they regarded them as a trend to a permanent state of affairs. The figure, however, clearly demonstrates that such

trends were accommodated within a normal distribution of events and so occurred by chance. Clearly, disaster of increasing magnitude or a recurring nature did not overtake them, as witnessed by their ability to continue to record the flood level at the Roda Gage, Cairo, for over a thousand years.

PETER H. ELLAWAY

Department of Physiology,
Charing Cross & Westminster
Medical School,
London W6 8RF, UK

1. Hurst, H.E. *Proc. Amer. Soc. Civil Engs*, 76, separate No. 11 (1950).
2. Jarvis, C.S. *Trans ASCE* Vol. 101 (1936).

Blind reviews

SIR—I was recently sent a manuscript to review (not from *Nature*) written by an individual whom I had never met, but whom I knew by (negative) reputation. I found the task of trying to ignore this knowledge in order objectively to review the manuscript to be difficult and disturbing. Recent discussions in the literature¹, in correspondence to *Nature* and at the First International Congress on Peer Review in Biomedical Publication (*JAMA* 263, 1311–1444; 1990) confirm that the issue of biased reviews is of general concern to the scientific community. Peer review of manuscripts is a critical step in the maintenance of quality and integrity in scientific communications. Reviews should be based solely on the merits of the research performed and the validity of the conclusions; the reputation, sex, race and institutional affiliations of the author(s) should not be allowed to bias the objective evaluation of manuscripts. However, the potential for bias is unavoidable unless reviews are blind, that is, the authors' names and institutional affiliations are removed before being sent to reviewers. Empirical studies^{2,3} indicate that reviewer bias can affect the publication of manuscripts, and that blind review does improve the quality of peer review. Blind review is commonly used in fields such as education, sociology, agricultural economics and psychology, but is uncommon in the physical, Earth and life sciences. In the interest of improving objectivity in the review process, I propose that blind review should be adopted as a general practice by scientific journals.

J. ELLEN MARSDEN

New York State College
of Agriculture and Life Sciences,
Cornell University,
Department of Natural Resources,
Fennell Hall,
Ithaca, New York 14853–3001, USA

1. Harnad, S. (ed.) *Behav. Brain Sci.* 5, 196–255 (1982).
2. McNutt, R.A., Evans, A.T., Fletcher, R.H. & Fletcher, S.W. *JAMA* 263, 1371–1376 (1990).
3. Peters, D.P. & Ceci, S.J. *Behav. Brain Sci.* 5, 187–195 (1982).

Strategic defence for the 1990s

Gregory Canavan and Edward Teller

Self-reliant defensive missiles for boost-phase interception seem to be cost-effective. Deployment would benefit all nations by decreasing the danger of rocket attack.

NOT surprisingly, political arguments and technical statements in the form of unproven assertions have clouded discussions of strategic defence. The following discussion is an attempt to provide a more objective basis by comparing the effectiveness of attack, defence, countermeasures and counter-countermeasures using various parameters and stating the essential assumptions on which comparisons are based. Our primary purpose is to clarify rather than to convince.

Paul Nitze proposed that it is worthwhile to deploy a defence if it is more cost-effective at the margin (disregarding the cost of research and initial deployment) than the countermeasures that could be used against it. That criterion for assessing novel defences has won considerable acceptance. We apply the Nitze criterion in a quantitative manner to nuclear attack missiles, to space-based defensive missiles designed to collide with the missiles as they leave the atmosphere, to anti-satellite missiles directed against space-based defences and, finally, to the costs of passive defensive countermeasures of hardening, evasive manoeuvre and decoys, singly and in combination.

The comparisons provide a coherent basis for assessing the effectiveness of the space-based defences now proposed for determining the most effective kind of defensive missiles. The most advantageous configuration can be specified more definitively because changing the assumptions about the capabilities of such defence changes effectiveness more than cost. Questions about the survivability of defence cannot be answered in absolute

terms, but the conclusions indicate how it can be bettered.

The Nitze criterion asserts the obvious truth that no military preparation is worthwhile if it can be countered at a lesser cost. Yet the criterion may miscarry because of the *fallacy of the last move*: the outcome may depend on where the chain of comparing moves and countermoves ends. The usual way to discuss the criterion is to propose space-based defensive missiles as a defensive measure and anti-satellite missiles as a response. We include attack, defence, countermeasures to defence, and passive measures against such countermeasures, and compare the effectiveness and costs of each in turn. Since the last point is discussed from the point of view both of the attacker and the defender, it provides a natural stopping point in the chain.

Present plans call for carrier vehicles ('carriers') housing some number of non-nuclear homing defensive missiles ('defenders'). That configuration and a newer design, a carrier with a single defender ('singlet'), are both assessed, as is the countermeasure of anti-satellite missiles ('anti-satellites').

The Nitze criterion cannot be applied rigorously and quantitatively when future technology is involved. Comparisons of cost have little meaning when the ratio is 2:1 and may only be indicative when the ratio is 5:1. But ratios of 20:1 can hardly be disregarded. Cost comparisons by themselves are not decisive, but trends of cost comparisons are more important.

Estimates of the cost of a missile are uncertain for a number of reasons, inclu-

ding the difficulties of comparing US expenditures and those of the Soviet Union. As a proxy, we base our comparisons of missiles and defenders on the masses to be lifted (payloads). They are also estimates, but their estimation is less uncertain. Money costs are roughly proportional to masses because launch costs are proportional to payloads. The cost of the colliding package depends on its mass, although construction costs will also depend on its complexity.

The main result of the discussion is that miniaturized, cheap and self-reliant space-based defenders are preferable to larger, more costly carriers. That result is not particularly surprising, but the detailed comparisons are illuminating and useful.

Attack missiles

The cost-effectiveness of attack missiles can be estimated from present missile costs. A typical Soviet missile, the SS-18 for example, has a payload mass of about 8 tonnes, including 10 weapons weighing in total about 3 tonnes, fuel weighing about 3 tonnes and structural arrangements weighing 2 tonnes¹. The US MX missile, of similar mass and payload, costs about \$200 million in similar silo-basing. That figure includes the roughly equal life-cycle operating cost and procurement cost. (Because those two cost components are typically comparable with each other, procurement costs might also be the basis for cost-effectiveness calculations, but we follow the conventional practice of using total costs.) If the costs of the US and Soviet missiles are similar, the present Soviet force of roughly 1,400 missiles has a total cost of about \$280,000 million.

Estimating the cost-effectiveness of nuclear-armed attack missiles is a logically necessary step. Thus, in a terrible sense, the cost-effectiveness of that offensive measure, which could destroy a nation with an estimated monetary value of \$10 million million, can be generalized to 36:1. (That gives attack a large advantage although overestimates of the effects of nuclear explosions may have caused many readers to guess a higher figure.) Of course, the effectiveness of the attack is diminished if defences are deployed.

A particular point is that missile attacks using conventional high-explosive warheads would not be cost-effective. Using conventional explosives would reduce the damage by a factor of 10^3 to 10^6 . That

Recent developments

Two changes have occurred in the interval between submission and publication of this Commentary. One is that the US Strategic Defense Initiative Office has adopted the small, singlet defensive vehicles called Brilliant Pebbles in preference to carrier vehicles housing several defensive missiles. Thus, portions of our discussion serve mainly to explain that decision. The second is the remarkable and hopeful change in Soviet-Free World relations. Yet, the rocket forces of the Soviet Union remain the strongest in the world by a large margin, and in spite of lessened tensions, further development of Soviet rocket forces seems to be continuing.

Even more significantly, the proliferation of rocket technology in the Third World shows no sign of abating, as the Iran-Iraqi war illustrates, and by the year 2000, more than twenty governments are likely to be so equipped. Fortunately, Brilliant Pebbles, while developed for strategic defence against attack from the Soviet Union, are apt to provide an even more complete defence against attacks from other quarters as well as against any accidental launch. The cost-effectiveness analysis provided on a bilateral basis applies equally to other situations, and the technical points presented here remain most relevant to international efforts to provide defence against aggressive use of rockets. □

Box 1. The absentee ratio

To reach a missile in boost phase, a defender must be within a distance equal to the product of its velocity and the time it takes the attack missile to accelerate and deploy its warheads. The interval, T , between launch and the end of warhead deployment for SS-18 missiles is about 600 seconds. If the effective radius of the area from which the missiles are launched is W and the defenders' velocity is V , a defender can reach a missile within a range $R = W + VT$ of the centre of the launch area. Taking W as 1,800 km and using the near-term value for V of 6 km s⁻¹, R is approximately 5,400 km, which encloses a geometrical fraction of $\pi R^2/4\pi R_e^2$ of the defensive constellation, where R_e is the Earth's radius. That fraction is approximately 0.18.

Concentrating the inclination of the defenders' orbits over the putative launch area has the effect of increasing the fraction of defenders within range by

a factor z of 1.1, giving an absentee ratio of about 20 per cent. This simple calculation agrees with near-exact analytical solutions within about 10–20 per cent.

If, in the mid-term, the attacker were to decrease each of W and T by a factor of two, geometrical availability would decrease by a factor of four. But for these smaller launch areas, the factor z would increase to about two, so that the overall absentee ratio would be about 10. Submarines could assemble before launch, but they could be addressed by absentee defenders.

The foregoing calculations neglect the circumstance that some of the re-entry vehicles will have been deployed before the defender arrives, which allows 20–40 per cent of the re-entry vehicles to escape. Increasing the number of defenders reduces that percentage and increases cost-effectiveness by a similar amount.

makes the connection between inter-continental missiles and nuclear weapons a quantitative proposition. We note in passing that the use of chemical or biological weapons at shorter ranges is susceptible to similar methods of analysis.

Space-based defence

The function of space-based defenders is to negate missiles by colliding with them. A general analysis would entail some complications, but the main issues can be discussed with simple models to the degree of accuracy we require.

Defenders and the sensors that support them would have to be predeployed at altitudes of 500 km or less in order to reach attack missiles before their weapons and decoys were released. Present defender designs have a total mass of about 100 kg, 90 per cent of which is propulsion. (Lighter defenders, which at present weigh about 40 kg, would reduce defender costs directly.) The carrier, which permits a number of space-based defenders to share common functions, could increase the total mass by 20–100 per cent, depending on its size and design^{3,4}.

If a 100-kg defender hit and destroyed a missile such as the SS-18 before separation of the warheads, it would destroy 6 tonnes of useful payload for an expenditure of 100–200 kg — a favourable mass-exchange ratio of about 60:1 or 30:1, depending on the weight of the carrier. A rough cost-exchange ratio can also be calculated, because 100–200-kg defenders have an estimated total cost of about \$6 million each. Therefore, the cost-exchange ratio would be about \$200 million: \$6 million or 33:1. (The cost of a current 40-kg defender is estimated at \$1.5 million making the cost-exchange ratio 133:1.) That confirms

that mass-exchange ratios are reasonable substitutes for cost-exchange ratios. However, the Soviet missile costs used are obtained indirectly, while the space-based defender costs are based on designs, not procurement costs, so that both have significant uncertainties.

The exchange ratios must be adjusted to take into account both the effectiveness of the defenders and the number of defenders within striking distance of a particular launch. The effectiveness of the defenders is high, so that a 10 per cent correction suffices.

Absenteeism is a more difficult problem. To be effective, a defender must be so placed as to reach an attack missile before it can deploy its warheads, which means that, in action against SS-18 boosters, only about 20 per cent of the defenders would be in a position to contribute to the defence against a simultaneous missile launch — an absentee ratio of 5 (Box 1 and ref.5). The remainder of the defenders would, of course, be available to counter staggered or dispersed launches, submarine-launched missiles and for defence against re-entering weapons.

The attacker could increase the absentee ratio of the defenders by replacing current missiles with missiles of shorter acceleration and deployment time ('fast-boost' or 'fast' missiles), which would reduce the proportion of defenders that could reach them. For full effect, fast missiles should be launched from a more compact area. The size of the launch area could be reduced by deploying fixed silos close together or by moving mobile launchers to within a few hundred kilometres of each other shortly before launch.

Reducing launch-time and radius by a half during the next decade would

increase the geometric absentee ratio four-fold, to about twenty⁶. If, however, the orbital inclinations of the defenders were chosen to concentrate them over missile launch areas, the fraction of defenders available would double, reducing the overall absentee ratio to about ten (see Box 1, refs 5, 9). Predictions about absentee ratios are obviously uncertain because of the unknown properties of future missiles and their basing.

Fixed missiles should be relatively cheap but could be less survivable. Mobile missiles should be more survivable, but could be individually more costly by a factor of between 2 and 10. Fixed missiles would support an offensive posture; mobile missiles could support either a survivable strategic withhold or a survivable countermeasure that could penetrate defences. Thus advanced missiles would probably be faster than present missiles, but more expensive, so in the exchange ratio, one attribute tends to cancel the effect of the other.

Estimating on the basis of current defender costs and correcting for absenteeism, the cost-exchange ratio would be \$200 million: \$30 million (\$6 million $\times$ 5), which is a little less than 7:1 in favour of the defence, and the exchange ratio in ten years (mid-term) would drop to little more than 3:1. While acceptable, both these figures could be improved by using smaller and thus cheaper defenders. The available technology is capable of producing defenders lighter by a factor of 3 than the designs we have assumed. Continuing efforts on weight reduction and other innovations may further improve the exchange ratio. Reducing costs by a factor of 3 would increase the near-term and mid-term exchange ratios in favour of the defence to 20:1 and 10:1.

Such advantageous ratios suggest that defenders should be effective in the near-term and mid-term, which contradicts the often repeated statement that defence may be offset simply by deploying more attack missiles. But all of that, of course, depends on the next link in the chain of argument — survivability.

Anti-satellite missiles

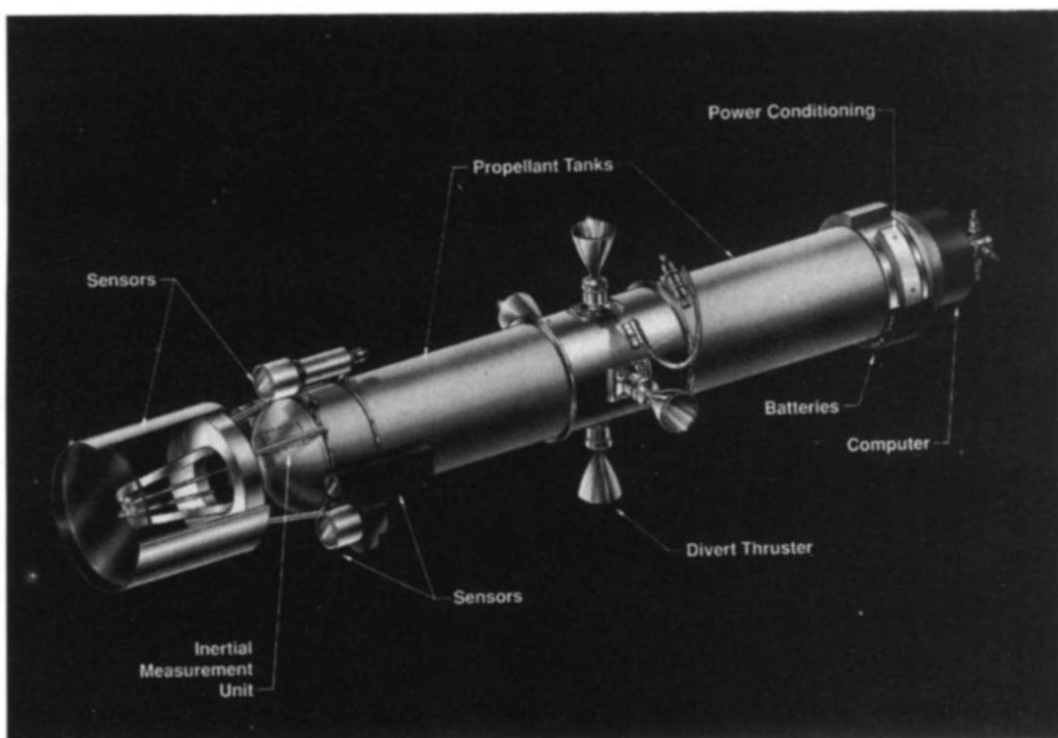
Throughout the next ten years, the most effective countermeasure to space-based defenders is likely to be a guided missile armed with a nuclear weapon. Non-nuclear anti-satellites would be lighter, but their decreased range of effectiveness would require a substantial increase in the accuracy and agility of anti-satellites if defenders were equipped with sensors to detect them and propulsions and decoys to evade them. Therefore, we shall assume that the Soviet Union will rely on its existing strength in nuclear technology. Weapons of 10–100 kiloton yield would be used to minimize damage to the launch area. Such anti-satellite weapons would

weigh about 400 kg, divided more or less evenly between the warhead and the guidance system.

Anti-satellites can use sub-orbital trajectories and thus need less final velocity than defenders, which must achieve orbit. If anti-satellites could use a vertical sounding trajectory, their weight could be reduced by a factor of 5 or 6. (But then a greater velocity would be needed to reach altitude and range before the arrival of the carrier and its dispersal of defenders.) The trajectory advantage of the anti-satellite is about a factor of 2, making its effective mass about 200 kg.

Estimating the mass of a defender at 100 kg and that of the rest of the carrier at 100 kg (a total of 200 kg) gives an exchange ratio of 1:1. But absenteeism reduces that ratio to 1:5 in the near-term and 1:10 in the mid-term. If the anti-satellite could intercept a carrier containing 10 defenders, the exchange ratio in favour of the anti-satellite would increase to 1:25 in the near-term and 1:50 in the mid-term if 100-kg defenders are employed. Present estimates of costs give an even more pronounced advantage to anti-satellites under those conditions: an anti-satellite attacking a 10-defender carrier of current designs estimated to cost \$2 million and \$60 million respectively would have cost-effectiveness of 30:1 in the near-term and 60:1 in the mid-term.

Those ratios are a strong argument for abandoning carrier vehicles with several defenders. The difficulties raised by anti-satellites make the necessity of further measures clear.



Artist's conception of a 1-metre-long Brilliant Pebble.

Passive protection

Passive measures for protecting defenders against anti-satellites, such as hardening, manoeuvrability and decoy, will engender obvious responses. Therefore, we shall discuss measures and countermeasures side by side. While hardening, manoeuvre and decoy cannot independently assure survivability, a combination of the three may succeed.

To assess cost-effectiveness, the masses required for hardening and manoeuvre are compared with the mass of the anti-satellite attacking the defence. Defence succeeds if the cost of the anti-satellite is higher, because the anti-satellite could then be effectively negated.

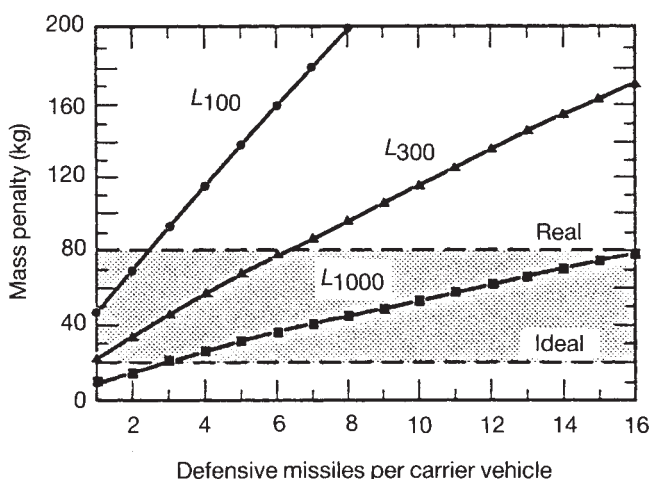
Early planning called for placing 1 to 10 defenders (n between 1 and 10) in a carrier

so that they could share the overheads for communications, sensors and other common requirements. Current plans emphasize singlets ($n = 1$), which have been called 'Brilliant Pebbles' (see the illustration above).

Our analysis shows that hardening by itself cannot assure survival against anti-satellites. The area of a carrier increases as $n^{2/3}$, and the hardening mass per unit area scales as the fluence J (see Box 2), which is proportional to $1/K^2$, where K is the distance between the target and the explosion. The total additional mass is proportional to $n^{2/3} K^2$. Without manoeuvrability, carriers on known orbits could be approached to within about 1 km by anti-satellites. For $n=10$ and $K = 10$ km, the penalty for hardening would be about 500 kg, which would not be acceptable. Moreover, there is no limit to anti-satellite accuracy, so there is none to the hardening penalty. Thus, hardening by itself cannot assure survival against anti-satellites.

Would the ability to manoeuvre provide a net advantage to the carriers? A carrier with 10 defenders of 100 kg each would weigh 1 tonne, and other structures could increase the total to as much as 2 tonnes, in which case (see Box 2) the fuel for manoeuvre would be 40 kg, representing a mass exchange ratio (400 kg: 40 kg) of 10:1 in favour of the carrier. After correcting for absenteeism (a factor of 10) and for anti-satellite trajectories (a factor of 2), both in favour of the anti-satellite, the exchange ratio is 1:2 against the defender. Even that result depends on the extent to which the anti-satellite's manoeuvres

FIG. 1 Penalty for hardening and manoeuvre. Relationship between mass penalty and number of missiles per carrier vehicle for different range to manoeuvre. The mass of the defensive missile is taken as 100 kg, as is the carrier overhead. The ideal anti-satellite mass is 20 kg, but is likely, in reality, to be 80 kg. The three lines on the diagram correspond to a range to manoeuvre of 1,000 km (■), 300 km (●) and 100 km (▲). Shaded area, probable effective mass of anti-satellite.



could offset those of the carrier.

What of a combination of hardening and manoeuvrability? Figure 1 illustrates the mass a hardened carrier would have to expend to protect itself as a function of the crucial range to manoeuvre (see Box 2) and the number of defenders per carrier. Comparing the expendable mass of the carrier, rather than its total mass, with the total mass of the anti-satellite (all of which is expended in the attack) is appropriate as long as the defence is successful. If the anti-satellite succeeds, the carrier would be destroyed, and its total mass would also be expended. The calculations of Figs 1 and 2 assume that the carrier survives.

In Fig. 1, the lower dashed line indicates the estimated effective mass (20 kg) of an ideal anti-satellite, and the upper dashed line the effective mass (80 kg) of an anti-satellite whose effectiveness is supposed degraded to 25 per cent by countermeasures and the strongly disturbed background produced by nuclear detona-

tions (see Box 2). The probable effective mass of an anti-satellite is represented by the shaded region.

Our estimate of the effectiveness (25 per cent) of an anti-satellite armed with nuclear weapons in attacking a hardened manoeuvrable defender may seem unduly low, especially in comparison with the high effectiveness we have attributed to the defenders. Two considerations are relevant. First, the booster leaves an extremely bright trail, while the defenders can be made inconspicuous. Second, electronics technology has been applied with excellent results to guidance systems — witness the successful use of Stinger missiles in Afghanistan — and is far more developed in the West than elsewhere. But with all that said, both in that and in less obvious connections throughout our argument, our assumptions are necessarily uncertain.

Our calculations show that the benefits of hardening and manoeuvre are separately

marginal, particularly against improved anti-satellites. Only hardened manoeuvrable carriers with a large warning range and a low mass lie within the region of probable effective anti-satellite masses shown in Fig. 1. The same calculations show that carriers become more survivable at all warning ranges as the mass of the defenders is reduced.

As seen in Fig. 1, the optimal conditions apply to singlet defenders ($n = 1$). Figure 2 shows the exchange ratios for singlets with masses of 100, 35 and 10 kg relative to a 20-kg anti-satellite when the distance (L) at which the defender receives warning and can begin to manoeuvre is 100–900 km. The first mass is typical of current defender designs, the second of a Brilliant Pebble based on the best recent technology and the third is a further step towards reduced weight that should be possible with techniques under development.

The calculations represented in Fig. 2 also show that the ratio of additional mass

Box 2. Carrier hardening, manoeuvre and decoy

Carrier vehicle hardening. A symmetrically exploding nuclear weapon of yield Y , attacking a defender hardened to withstand fluence (radiation flux per unit area) J , has an effective range K equal to $(Y/4\pi J)^{1/2}$. Satellites are at present protected against fluences of a few joules per cm^2 and so would be destroyed by a 10-kilotonne detonation at 10 km. Hardening with a few grams per cm^2 of shielding allows a reentry vehicle to withstand the radiation flux of a megatonne burst at 10 km (about 300 joules per cm^2 (ref. 9)).

That amount of radiation does not destroy the reentry vehicle, but is assumed to crack the reentry heat shield. Defenders need not reenter, so the integrity of their shields is not essential, which makes it easier to assure their survivability. Assuming that the amount of shielding needed is proportional to mass, about 1 kg m^{-2} of shielding should suffice to protect a defender against a 10-kilotonne attack at $K = 10 \text{ km}$, but our calculations use 10 kg m^{-2} .

Carrier vehicle manoeuvre. To avoid interception by an anti-satellite of effective range K , a carrier would have to manoeuvre through a greater distance than 10 km. The distance at which the carrier receives warning and can initiate its manoeuvre ("range to manoeuvre" or L) is crucial. If the carrier is warned when the anti-satellite is at a range $L = 1,000 \text{ km}$, the required angular deflection (K/L) would be given by $\delta V/V$, where V is the orbital velocity (about 7 km s^{-1}). With fuel of specific impulse c (about 3.5

km s^{-1}), the fractional mass expended is $\delta M/M$, which is approximately $\delta V/c = (\delta V/V)(V/c)$ or, roughly, $2K/L$. That means that a carrier of mass M would have to expend fuel approximating $2M(K/L)$ in an evasive manoeuvre.

Combined hardening and manoeuvre. Both the penalty for hardening ($n^{2/3}/K^2$) and the penalty for manoeuvre ($2M(K/L)$) favour carriers with fewer defenders per vehicle. Their sum is minimized by choosing an effective range K proportional to $(n^{2/3}L/M)^{1/3}$. Hardening and manoeuvre combined would then have a minimum penalty of additional mass proportional to $(n^{1/3}M/L)^{2/3}$. Scaling both on the basis of number of defenders per carrier ($n^{2/3}$) and on the basis of mass $M^{2/3}$ (which increases with n) indicates that a carrier with one small defender is advantageous. Scaling on the basis of the range to manoeuvre ($L^{-2/3}$) makes it clear that extending that range by means of countermeasures, jamming or signature cancellation is beneficial to defence.

With range to manoeuvre of 1,000 km, a singlet would need an additional 10 kg (3 kg for hardening and 7 kg for manoeuvrability, which proportion holds for various parameters). That gives a singlet a 2:1 advantage over an ideal anti-satellite (with an assumed effectiveness of 100 per cent). An advantage persists up to carriers of four defenders ($n = 4$).

Decreasing the range to manoeuvre to 300 km increases the additional mass required by a singlet to about 20 kg, which reduces the mass-exchange

ratio to unity against ideal anti-satellites and to about 6 against real anti-satellites (with an assumed effectiveness of 25 per cent). Reducing the range to 100 km puts singlets at a disadvantage against ideal anti-satellites, although they still retain about a twofold advantage over real anti-satellites.

The actual range to manoeuvre is determined by the ability of the defenders to detect and identify the bunching of an anti-satellite. That depends on future developments and deployments, so the value of L cannot be specified. The best that can be done is to assess its impact parametrically.

Decoys. The following considerations would apply if a carrier, on detecting the approach of an anti-satellite, were to eject a number of decoys that would drift to distances of a few times the accuracy range (K) of the anti-satellites, in the approximately 100 seconds before the anti-satellite arrives.

Without sophisticated sensors, the anti-satellite would have to pick an object and attack. For N decoys, the carrier would have a probability of about $1-1/N$ of surviving the attack. But an anti-satellite might have optical or infrared sensors, so decoys would have to have the right shape and temperature. Decoys would also need some strength, or the anti-satellite could use one burst to destroy the decoys and a second to attack the carrier. Credible decoys would probably require masses of about 1 kg.

expended in manoeuvre by a 100-kg hardened singlet with $L = 900$ km to the mass of an ideal anti-satellite is approximately 3:1 in favour of the singlet. Against anti-satellites that could reduce the range to manoeuvre to about 100 km, the exchange ratio would be below 1. With $L = 900$ km, a 35-kg singlet would have an exchange ratio of about 6, but that again would drop to unity if the anti-satellite reduced the range to 100 km. With $L = 900$ km, a 10-kg singlet would have an exchange ratio of 13:1; with a range to manoeuvre of 100 km, the exchange ratio would still be 3.

Thus, lightweight singlets appear survivable, provided the cost-estimates are valid. But costs can be established only with production runs of about 100 units. The deployment of 100 lightweight defenders could provide effective protection against accidental or third-country launches, but 10 to 100 times that number would be needed to provide significant

For decoys to drift to a distance a few times K would require velocities of about 0.3 km s^{-1} , a small movement difficult to observe. Eliminating decoys separated by distances of two K or more would require the attacker to commit an anti-satellite to each object, implying an extra 20 kg per decoy.

Deploying decoys in a rough sphere would minimize the mass required. To make all locations in the sphere credible, the carrier would have to be able to move to any part of it. For N decoys, the radius of the sphere is $N^{1/3}K$, so the fuel for repositioning would have to be increased by $N^{1/3}$. If 30 decoys were used, the fuel needed for repositioning would be tripled.

Figure 3 shows the cost-effectiveness of various numbers of decoys (1 kg mass) for carriers (weighing 100 kg) with 1–9 defenders (100 kg mass), where the range to manoeuvre is 1,000 km. The ordinate of the figure reports the ratio of the effective mass of ideal (20 kg) anti-satellites needed to attack a carrier vehicle relative to the extra mass expended for decoys and escape. The cost-effectiveness ratios shown in Fig. 3 are less than those in Figs 1 and 2 because it has been assumed that when an attacker commits as many weapons as there are defensive objects, the carrier vehicle would also be destroyed, whereas the previous calculations were on the basis that the carrier would always survive. That has the consequence that for zero decoys, the carrier is destroyed and the cost-effectiveness is between 0.1 and 0.01.

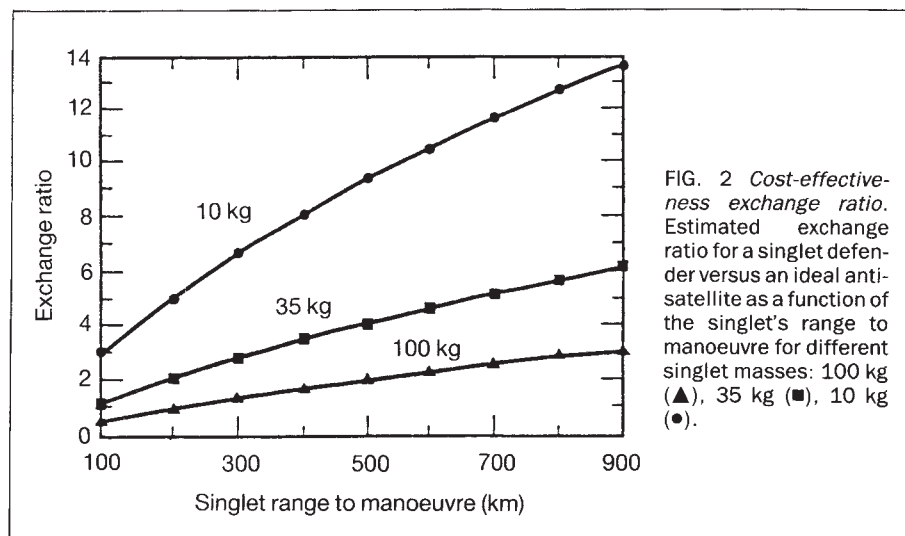


FIG. 2 Cost-effective exchange ratio. Estimated exchange ratio for a singlet defender versus an ideal anti-satellite as a function of the singlet's range to manoeuvre for different singlet masses: 100 kg (▲), 35 kg (■), 10 kg (●).

boost-phase defence.

The survivability of defence would also be enhanced by increasing the number of objects anti-satellites must attack. Hardening and manoeuvre in combination may allow lightweight singlets to be cost-effective, but as will be seen the use of decoys can establish by itself a substantial advantage for the defence.

The calculations summarized in Fig. 3 (see Box 2) show the additional advantages that can be gained by using decoys. Adding 20 decoys to a singlet increases its mass by 36 kg (20 kg of decoys and 16 kg of fuel and shielding). To counter that, the attacker would need 21 anti-satellites of effective mass 20 kg each. If the attacker committed a weapon to each object around the carrier, the defenders would all be destroyed, but the penalty on defence would then be $36 \text{ kg} + 200 \text{ kg}$ (the mass of the singlet) and the exchange ratio (420:236) about 1.8.

For a singlet with 100 decoys, the extra mass totals 323 kg, and the overall ratio is 2,020:323, or about 6.3:1, which is a significant improvement. For 200 decoys, the exchange ratio for a singlet is about 9.4. Decoys can also help carriers, but not as markedly; a carrier with 9 defenders and 200 decoys would require 128 kg of additional fuel and shielding, a total of 1,328 kg (200 + 128 + 1,000). The exchange ratio $(20 \times 201): 1,328$ is about 3, which is less than half the exchange ratio for a singlet with 200 decoys. A singlet with 400 decoys has an exchange ratio of about 13:1.

Thus, a combination of hardening, manoeuvrability and decoys is three times more cost-effective than hardening and manoeuvre combined.

The addition of each decoy requires the attacker to add a 20-kg weapon, which produces a limiting exchange ratio of 20:1 in favour of the defence. To reduce the exchange ratio, the attacker would have to discriminate the decoys in a short time at a great distance, a difficult task for which the defence has further responses. More-

over, the contest would be conducted in a field of technology in which the West has proven capabilities.

At present, the only other countermeasure foreseen that could alter the situation in the mid-term is that of active defence — the destruction of anti-satellites. That turns out to be marginally useful, but would make it necessary also to take account of the countermeasures available to the anti-satellite. We shall not pursue those questions here.

Thus we have considered the effect of the passive measures available to the defence from the point of view of both defence and attack. The discussion of moves and countermoves ends at this point. Hardening, evasion and decoy in combination appear to provide a significant advantage to defence in a cost-effective manner.

Vulnerability in general

Further countermeasures to frustrate the defence have been suggested. Estimates of their cost-effectiveness are not appropriate at this time. What follows is a necessarily qualitative consideration of four such possibilities.

Beam weapons. At present, no nation has a beam weapon deployed in space, although the Soviet Union has some ground-based beam weapons that could become effective against a space-based defender, and particularly against carriers with several defenders. Even an expensive beam weapon could be cost-effective in countering space-based defences if it could protect a dozen or more attack missiles moving on roughly parallel paths.

The natural countermeasures against such a weapon are near-invisibility and great numbers of decoys, each of which becomes the more feasible if small independent space-based defensive missiles are used. Synthetic construction materials would provide a smaller signature, and absorbing paint would help to frustrate visual observation. In addition, small corner reflectors unfolded at some

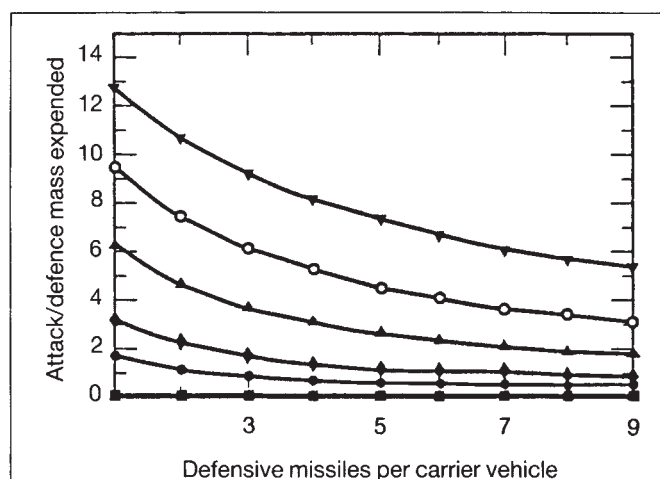


FIG. 3. Cost-effectiveness exchange ratio with decoys. Estimated exchange ratio for a defender with decoys versus an ideal anti-satellite as a function of the number of defensive missiles per carrier vehicle. The mass of the defensive carrier is taken as 100 kg, as is the carrier overhead; the mass of a decoy is taken as 1 kg, the range to manoeuvre as 1,000 km and the anti-satellite mass as 20 kg. The curves show the exchange ratio for different numbers of decoys per carrier: no decoys (■), 20 (●), 40 (◆), 100 (▲), 200 (○), 400 (▼).

distance from the defender could serve as sensor decoys. Suppressing the visible and infrared emission from the jets propelling the defenders seems necessary and may be the most difficult to achieve.

Red-out from nuclear explosions. The detonation of a few hundred nuclear explosives in the high atmosphere would produce sufficient heat and radiation to make the launch of attack missiles and anti-satellites difficult to observe. (Such explosions would be virtually harmless on the surface of the Earth.) The condition, called red-out, would precede the attack, and the explosions would occur 2,000–3,000 miles from the launch sites, essentially over the high seas.

Such an attack strategy would not be significant because the nuclear explosions would give the defenders adequate warning to take evasive action. Concealing the plumes of the missiles in enhanced background could be more effective, although the strength of their signatures makes that unlikely.

Attrition during peacetime. Another scenario would employ countermeasures against space-based defenders during peacetime. While anti-satellites would be effective in that role, their release might be viewed as an attack, so the main danger in this case would probably come from ground-based lasers. Satellites might arguably disturb the aim or phase control of the laser beam, but such action against instruments in a foreign country could result in an even more serious form of hostile exchange.

The ideal response would be to deploy defenders so inexpensive that their replacement costs would be less than the

cost of destroying them. That is the essential goal of Brilliant Pebbles and is one of the reasons that space-based defences have evolved from large carrier vehicles to singlets.

Vulnerability from complexity. A complex design is vulnerable in each of its components and in every link between them. Space-based defender designs that depend on sensor satellites and carrier vehicles in addition to the defenders which they can eject are much more vulnerable to failure. Because the success of the defence rests on the weakest link, the

Conclusions

This study has two quite different conclusions, one probable and the other almost certain. Space-based defences using miniaturization, sensors, computers and decoys appear to have a reasonable chance of providing significant defence in the near future. That conclusion is, of necessity, uncertain. Too many assumptions must be made for definitive quantitative conclusions to be possible. Of particular importance is the cost of singlets.

The second conclusion is more certain. For survivability, defenders are best deployed in self-reliant singlets. In 1983, the singlet proposal would have evoked scepticism: today, Brilliant Pebbles are clearly feasible. Practically every aspect of this study points to the conclusion that singlets are the best line of deployment. The US Strategic Defense Initiative Office has recognized the advantages and decided to develop singlets in preference to large carrier vehicles.

The cost-effectiveness studies in this paper lead to a practical recommendation about decoys. If the feasible countermeasure of an anti-satellite missile were adopted, decoys released in the last few minutes of its approach would appear to be decisive for the success of defence. Therefore, considerable research and development effort should be given to decoys, particularly to the question of how

the weight of a sufficiently imitative decoy can be reduced.

The effectiveness of decoys could be improved by constructing singlets with superficial but highly noticeable differences so that the decoys need not imitate a single, well-defined signature. Singlet construction could also use modern composites rather than metals to minimize radar reflection.

This study also makes clear the difficulties of making decisions about defence, given the uncertainties about its costs. Deployment of 100 singlets (without decoys) would permit more accurate information about the function and cost of a substantial system. It would also provide a good (though never complete) defence against missiles fired accidentally or by a terrorist state. (Unfortunately, the availability of missiles that could carry any of several highly effective tools of terror is steadily increasing.) Such deployment would mean that for the first time since rockets armed with nuclear warheads were deployed a quarter of a century ago, people within range of those terrible weapons would not be completely undefended against them.

The cost of deploying 100 Brilliant Pebbles is probably less than \$1,000 million. Such deployment is an urgent and necessary step with several benefits. Political objections to a pilot development project could be overcome by making the enterprise an open and international undertaking. That would be particularly fitting because, since its inception, the Strategic Defense Initiative has been intended to protect the whole of mankind, not just a single nation. A cooperative demonstration supported in part by contributions from other nations would be a highly constructive step towards assured safety for all. □

Gregory Canavan is scientific advisor for defence programmes at the Los Alamos National Laboratory, and Edward Teller is director emeritus of the Lawrence Livermore National Laboratory, and a senior research fellow at the Hoover Institution, Stanford University, Stanford, California 94305–6010, USA.

1. Bloembergen, N. & Patel, C. *Rev. mod. Phys.* **59**, S28–S30 (1987).
2. Office of Technology Assessment, US Congress, *MX Missile Basing*, OTA-ISC-140 (GPO, Washington, DC, 1981).
3. Gardner, J., Gerry, E., Jastrow, R., Nierenberg, W. & Seitz, F. *Missile Defence in the 1990s* (Marshall Inst., Washington, DC, 1989).
4. Canavan, G. *Defense Platform Size and Survivability*, Los Alamos National Laboratory report LA-11244-MS (1988).
5. Canavan, G. & Petschek, A. *Concentration of Defensive Satellite Constellations*, Los Alamos National Laboratory report LAUR-3510 (1988).
6. Canavan, G. *Directed Energy Architectures*, Los Alamos National Laboratory report LA-11285-M5 (1988).
7. Canavan, G. in *SDI: The First Five Years* (conference proceedings) (Washington, DC, 1988).
8. Garwin, R. *Nature* **315**, 286–290 (1985).
9. Garwin, R. & Bethe, H. *Sci. Am.* **218**(3) 21–31 (1968).

ACKNOWLEDGEMENTS. We thank Lowell Wood for a number of stimulating discussions.

Complicated measures of complexity

The search for a means of telling the complexity of numerical data has been urgent but frustrating. Now, there may be some relief in the assertion that there is no single measure.

EVERYBODY wants to be able to measure complexity, but nobody quite knows how it should be done. That is the impression one is likely to form by skimming through the literature on the subject, which is understandably spilling out in the most unexpected places — mathematics journals, those concerned with computer science and communications and straightforward physics journals as well as, inevitably, in Wolfram's *Complex Systems*.

The issue has evidently been made pointed by the recent wave of interest in deterministic chaos. But why is it so difficult?

The essence of the problem is neatly put by Peter Grassberger of the University of Wuppertal (*Int. J. theor. Phys.* **25**, 907; 1986) in these terms. Imagine three two-dimensional patterns, one of them an array of black and white squares as on a chessboard, one the now-familiar kind of trajectory traced out in two dimensions by a chaotic system — typically an intricate pattern of nearly periodic orbits grouped together into whorling shapes — and the third simply a random pattern of dots. Both the first and the third are inherently simple. A few simple rules will help one to construct a checkerboard, while a random pattern may need even fewer rules. The chaotic pattern, on the other hand, can be reproduced only by simulating the system that generates it. Intuitively, it is by far the most complex. But this is not what the numbers say, or at least some of them.

To the extent that many measures of complexity boil down to specifying the information content of a pattern, they turn out to be the equivalents of entropy. And entropy, of course, is greatest for disordered systems. The entropy of the random pattern is thus greater than that of the chaotic pattern, which in turn is greater than that of the checkerboard. Such measures of complexity, including those derived from Shannon's theory of communications systems, all give pride of place to the random pattern of dots, which offends expectation.

The standard answer may be counter-intuitive, but the antecedents of this question are interesting in themselves. To make progress, one needs a way of generating patterns with more or less complexity. To make life simple, it is best to begin with one-dimensional patterns when, 'without loss of generality' as they say, one can conveniently think of a pattern as consisting of an infinite string of symbols which may, to suit the computers,

be either '0' or '1'. To begin with, one needs a way of generating such patterns.

Grassberger's article turns out to be a neat and intriguing review of much that has been done to define measures of complexity. It is intriguing that he quotes two essays in biology — an attempt to define the most economical taxonomic tree, and to describe life mathematically — as part of the inspiration of his own account. (The second of these authors, G. J. Chaitin, is cited as the source of an interesting view on the meaning of randomness in a piece of Scientific Correspondence we shall be publishing next week.)

Constructing symbol strings with simple structure is easy. By alternating '0's and '1's, one generates simple strings such as '01010101...'. Generating more complex patterns requires more ingenuity — one might wish to generate patterns in which the sequence '...111...' never appears, for example. That, it turns out, can be done by constructing an appropriate directed graph, which is nothing more than a set of points (nodes) on a plane which are joined to their neighbours by lines which carry arrows to show the direction in which movement is permitted. Each line is also marked by one of the symbols making up the string, either '0' or '1'. The rule is to find the starting point (which may have only two arrows leading away from it), to flip a coin at each node at which one arrives and to take as the next symbol in the growing string the symbol on the line chosen as a consequence.

Take a simple square, for example, with all the arrows following in the same direction, but with '0's and '1's alternating around the square. The sequence generated is simply '010101...' with an appropriate choice of corner at which to start. But if one adds a single diagonal from the same corner, but carrying the symbol '1', sequences such as '0101' and '101' will be mixed in the string at random. Much more complicated patterns can easily be generated. The connection with chaotic systems can be made direct by using a generating function, of the kind that chaos-mongers live and breathe — something such as $x_{i+1} = f(x_i)$. The idea is that this equation is a way of indefinitely remapping the points in some interval of the x axis onto the same straight line.

The practitioners in chaos are usually more interested in mapping some finite interval — say that between 0 and 1 — onto itself, which is convenient for pattern generation because it is then possible to

assign to x_{i+1} the value '0' if it is less than 0.5 and the value '1' if it is 0.5 or greater.

One of the most intriguing features that keeps cropping up in this literature of pattern formation and analysis is that there are repeated references to Noam Chomsky, who is largely responsible for putting the grammar of language on a logical basis. The reason is straightforward — strings of two symbols with restrictions, such as '...111...' is disallowed' can be linked to strings of letters, or words, and the restrictions to the rules of grammar. Indeed the set of all the algorithms by which the patterns are generated becomes the set of all the rules of an infinite set of grammars. In language, the absorbing question is the degree to which rules like these represent real language.

That may seem a diversion, but it is an important one. Another is with the theory of computation and the Turing machine, which is a way of relating the complexity of a computational task to the characteristics of the ideal computer required to carry it out.

Grassberger, for what it is worth, defines four measures of complexity for strings like these, of which he says he believes the most relevant for physical problems is that called *effective measure complexity*. He goes on to show that the definition allows for the complexity to be infinite which the entropy is zero.

But now, it seems, a more radical view has come to the surface. G. d'Alessandro and A. Politi of the Istituto Nazionale di Ottica at Florence have made a direct attack on the difficulty that random patterns are assigned great numerical complexity by a quite novel procedure (*Phys. Rev. Lett.* **64**, 1609; 1990).

What they imagine is that the complexity of a string will eventually be measured by a hierarchy of numbers, the first of which is akin to an entropy and defined as the logarithm of the number of admissible sequences (given the grammar) of length n divided by n as the numbers become very large. But why not do the same for the forbidden words, deriving a second measure in exactly the same way? Out of that definition tumbles directly the notion that the complexity of a random string is identically zero. Their next task is to generalize the argument to several dimensions. One is naturally left wondering why it should ever have been thought that there could be a single number measuring complexity of very different kinds.

John Maddox

Quest for magnetic monopoles

Henry J. Frisch

THE great appeal of magnetic monopoles is that their existence would both explain why electric charge is quantized and provide clues about the structure of the Universe moments after the Big Bang. Furthermore, whole classes of theories in particle physics work only if they exist. All of which explains why two groups, one from Stanford and the other a collaboration between IBM and Brookhaven National Laboratory, have spent the past three years searching for these elusive objects^{1,2} — without success — and look forward to doing so for years to come.

Most hypothetical particles, such as squarks, wimps, winos and technipions, may be loved by their mothers but are easy for the rest of us to shrug off. Given that magnetic monopoles must exist according to theories unifying the four fundamental forces — grand unified theories, or GUTS — they are less easy to ignore. One expects, however, only one monopole per universe on average in such theories, and there is no *a priori* reason to believe that the number of monopoles will be any more than this. In particular, the presence of galactic magnetic fields sets an upper limit on the number of monopoles, as they would be accelerated and would thus consume the fields by extracting energy from them — this limit is known as the Parker bound³.

Monopoles are predicted to be very heavy with a mass comparable to the energy believed to be necessary for the unification of the four forces, namely 10^{16} times the mass of a proton, indeed more like the mass of a large bacterium. Consequently, monopoles are expected to move

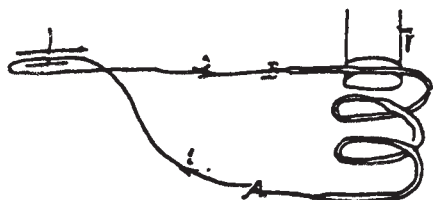


FIG. 1 Faraday's sketch of his experiment. The induced current made a compass needle swing during the magnet motion.

slowly (at least to a particle physicist) with respect to the Earth, with a relative velocity of $1/1,000$ the speed of light. In essence, we are sweeping through a sea of cold monopoles, if indeed there are any left over from the Big Bang.

The Stanford and IBM-BNL groups have set forth^{1,2} new limits on the flux of magnetic monopoles in cosmic rays. The instruments used are highly sophisticated descendants of an apparatus for detecting magnetic monopoles in cosmic rays that started with Luis Alvarez⁴ and that goes back to the very first experiments on

magnetic induction by Michael Faraday⁵.

The Stanford and IBM-BNL experiments both work on the principle used by Faraday in his first detection of magnetic induction. In his laboratory notebook on 8 December 1831, Faraday wrote "When the marked pole of the magnet entered the helix as in the figure [Fig. 1] the marked end of the needle went WEST, hence electrical currents as indicated by the arrows. On withdrawing the magnet, the MARKED END went EAST". Faraday's pole was one end of a dipole magnet, and as the magnet was long, his inserting the pole through the loop had exactly the same effect as a magnetic monopole traversing the loop. (If one cuts a dipole magnet in half, one has two dipole magnets rather than monopoles, as magnetism as we know it comes only from electric currents. It is like the little boy who cut a worm in half, saying "there, now you have a friend".)

The Stanford group, led by Blas Cabrera, has built a detector consisting of an array of loops of niobium wire arranged as the sides of a 5-m-long octagonal solid, with superconducting quantum interference devices (SQUIDS) as current detectors. A monopole passing through the apparatus will penetrate two sides, inducing current in each of two loops (this is just the effect that Michael Faraday was describing above). The loops are folded into subloops in such a way that the field from neighbouring subloops is of the opposite sign. This folding makes the stray fields from the loop fall off rapidly with distance from the plane of the loop, and makes the mutually inductive coupling to a surrounding superconducting shield very small^{6,7}. By combining subloops in either series or parallel, the inductance of the loop is made the same as the input impedance of the SQUID⁸. The whole arrangement, loops and shield, is cooled to liquid helium temperatures by a closed-cycle helium liquefier. The Stanford detector, which has an effective area to cosmic monopoles of 1.1 m^2 , has now had a total exposure time of 6,482 hours with no candidate events. This corresponds to a flux limit of less than $7.2 \times 10^{-13} \text{ cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}$ at 90 per cent confidence level, as shown in Fig. 2.

This is the third-generation detector that Cabrera and his group have built. The first, a technological *tour de force*, consisted of a 5-cm-diameter loop in an ambient field of 5×10^{-8} gauss (the new detectors operate on the principle⁷ that the ambient field is held constant by superconducting shields, rather than made tiny by herculean efforts as in the first-generation detector of Cabrera's). This

first experiment created a flurry of excitement when it recorded a single candidate event⁹. The experiment would have seen 2,000 events by now if the one event had been real. It is now believed¹⁰ that the one event was caused by motion of pinned magnetic flux in the detector, and was not a monopole. (Those in the field, however, joke that it may have been the only monopole in the Universe, and having traversed California, it's not coming back soon.)

The IBM-BNL experiment is similar in concept to that of the Stanford group. The subloops are made on printed circuit boards in a more complicated pattern, and the loops are arranged in a rectangular, rather than octagonal solid. Perhaps the biggest difference is that the IBM group uses SQUIDS designed and made in-house in a production line. The experiment, which has an effective sensitive area of 1.0 m^2 , has run for longer than the Stanford detector, 13,410 hours, and has seen no candidate events. The flux limit deduced from this experiment is $3.8 \times 10^{-13} \text{ cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}$, and when the two experiments are combined, the upper limit is calculated to be $2.2 \times 10^{-13} \text{ cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}$, a factor of 6,500 below the 'flux' limit given by the one erstwhile candidate of 1982.

The significance of these results is twofold. The experiments have conclusively ruled out theories in which monopoles give and take energy from the magnetic field in an oscillatory way. But perhaps more importantly, they are a step on the road to making detectors with effective areas in the $1,000 \text{ m}^2$ range¹¹. The new high-temperature superconductors, even with their limited current-carrying capability, are ideally suited to making very large arrays of loops¹². And the new superconductors do not require cryogenic dewars or fancy SQUIDS for readout, but can be used with ordinary junction-

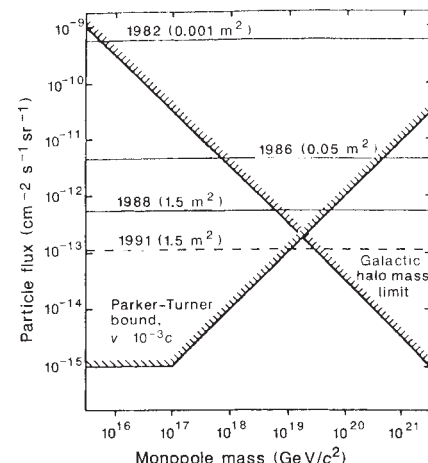


FIG. 2 Comparison of experimental and theoretical limits on the flux of magnetic monopoles in cosmic rays. The limits from induction experiments hold for all monopole masses, and hence are horizontal lines. Astrophysical limits depend on mass and are shown as hatched lines (from ref. 1, slightly modified).

field-effect transistors¹².

Other, bigger detectors are searching for monopoles using the quite different technique of detecting their ionization of matter¹³. But the Faraday technique, of inducing a current in a coil by the passage of a magnetic pole, is the only one that is exactly calculable, and also the only one that measures both the sign and magnitude of the magnetic charge. It is, moreover, a technique that can be calibrated by putting a magnetic pole (one end of a long dipole) directly through the detector.

Why do physicists spend years searching for monopoles in the face of long odds? Monopoles, if found, represent high-energy physics at energies of (probably) 10^{16} gigaelectronvolts, far beyond the 10^4 gigaelectronvolts of the Superconducting Super Collider being built in the United States. The monopole mass is a measure of the energy at which we believe the four fundamental forces become unified. It is only by looking at the relics of the Big Bang that we can explore the energies corresponding to the earliest moments of the Universe. These physicists are working at the highest energies of particle physics, but in small experimental groups, and in their own laboratories. The experimenter deals with arcane theory, that of the grand unified theories, but uses classical electromagnetism in its purest form in the experiments (we feel Faraday and Ampère watching over our shoulders).

From these experiments and the opportunities offered by the high-temperature superconductors, the path is clear to build a really big monopole detector, with an effective area of say, $1,000 \text{ m}^2$. It would be a long shot, and would probably just lower the limits on the monopole flux to the level of the Parker bound. But there remains a deep longing to see the trail of flux left by what we believe is the heaviest fundamental particle remaining from the Big Bang — and even though we believe that they are terribly rare, one can always hope to get lucky. □

Henry J. Frisch is in the Physics Department and the Enrico Fermi Institute, University of Chicago, 5640 Ellis Avenue, Chicago, Illinois 60637-1433, USA.

Viral hijack of receptors

Elliott M. Ross

COMMUNICATION systems are vulnerable to subversion, and cellular communication is no exception: G protein-mediated signalling systems provide targets for disruption of whole networks. These tripartite systems are composed of plasma membrane-spanning receptors that bind hormones or neurotransmitters; GTP-binding proteins (G proteins) that integrate and amplify signals from several receptors; and effector proteins — enzymes, ion channels or transporters — that generate intracellular second messengers such as cyclic AMP. Bacteria can attack this type of signalling system with toxins that either block G protein function or stimulate overproduction of second messengers (mechanisms manifest, for example, in cholera and whooping cough respectively). The mastoparans, insect toxins, masquerade as G protein-coupled receptors and thereby activate the G proteins, while some oncogenes encode aberrant or inappropriate G proteins and receptors¹.

A novel subversion of G protein-mediated signalling is illustrated in a report on page 774 of this issue by Mark Chee and co-workers² in Bart Barrell's group at the MRC Laboratory of Molecular Biology at Cambridge. They describe the presence of three genes in the genome of human cytomegalovirus (CMV) that appear to encode G protein-coupled receptors. Human cytomegalovirus, a member of the herpes virus group, is a widespread, often sexually transmitted pathogen. If expressed, these receptors might influence the course of viral infection, the choice of the cells that become infected, or the way in which infected cells become sick.

The three receptor genes were identified in a scan of open reading frames (ORFs, DNA sequences likely to encode proteins) in the genome of human CMV, which has been completely sequenced by Barrell's group³. Two were identified in 1986⁴ as probably encoding membrane proteins and the third is now described here. The authors could predict a function for the products of these ORFs because the structure of the G protein-coupled receptors is highly conserved in this family of homologous proteins. They all contain seven stretches of relatively hydrophobic amino acids that span the plasma membrane, with the amino terminus on the extracellular face and the carboxyl terminus in the cytoplasm. Even when overall sequence identity is slight, the overall structure, a few conserved amino acid sequences, and the positions of about ten invariant amino acids are maintained. The proteins encoded by the three CMV ORFs have all of these features. Other unrelated seven-span pro-

teins, such as hydroxymethyl glutaryl-CoA reductase and bacteriorhodopsin, do not meet these criteria. Neither do several other putative seven-span proteins encoded elsewhere in the CMV genome. Therefore, although it is risky to assign a specific function to a protein from its DNA sequence, these predictions are reasonable.

It is easy to imagine that CMV acquired the genes for G protein-coupled receptors from the genomes of host cells, much as tumour viruses acquire DNA sequences that function as oncogenes. Chee *et al.* have not yet shown that their ORFs encode active receptors or that they are ever expressed, but they do point out that the three ORFs are found in the 'unique' region of the viral genome. This region is made up of relatively densely packed ORFs that encode functional proteins expressed at distinct points in the CMV life cycle³. Nevertheless, it will be important for virologists to determine if and when these novel genes are expressed.

Finding out how these receptors function will be challenging. The G protein-coupled receptors number over 100, and the ligands that regulate the CMV-encoded receptors are unknown, but a few structural motifs may provide some clues. The short cytoplasmic loops that connect spans 5 and 6 in the CMV-encoded receptors are characteristic of the receptors for the peptide hormones substance P and substance K, and of the *mas* oncogene product, which is probably also a peptide receptor. The CMV-encoded receptors might thus also be peptide receptors. They do not appear to be opsins or amine receptors (like the adrenergic, serotonergic or muscarinic receptors, for example) and they lack the large amino-terminal domains that characterize the receptors for thyroid-stimulating and luteinizing hormones. On the other hand, some viral oncogenes that are descendants of cellular receptors have often lost their ligand-binding components. It is possible that the CMV-encoded receptors might have lost their capacity to bind the ligands of their evolutionary precursors and are constitutively able to activate G proteins.

The receptors' G protein targets may also be difficult to identify. There are now nearly twenty identified mammalian G proteins that regulate about ten ion channels and perhaps as many enzymes. A receptor's selectivity among G proteins seems to be determined by sequences in the second cytoplasmic loop and the amino-terminal portion of the third cytoplasmic loop, but our ability to identify the G protein targets of a receptor solely on the basis of its sequence is still rudimen-

1. Huber, M.E. *et al.* *Phys. Rev. Lett.* **64**, 835–838 (1990).
2. Bermon, S. *et al.* *Phys. Rev. Lett.* **64**, 839–842 (1990).
3. Turner, M.S., Parker, E.N. & Bogdan, T.J. *Phys. Rev. D* **26**, 1296–1305 (1982).
4. Alvarez, L.W. *Lawrence Radiation Laboratory Physics Note No. 470* (1963).
5. Martin, T. (ed.) *Faraday's Diary Vol. I* (Bell, London, 1932).
6. Incandela, J. *et al.* *Phys. Rev. Lett.* **53**, 2067–2070 (1984).
7. Tesche, C.D. *et al.* *Appl. Phys. Lett.* **43**, 384–386 (1983).
8. Somalwar, S., Frisch, H., Incandela, J. & Kuchnir, M. *Nucl. Instrum. Meth.* **226**, 341–343 (1984).
9. Cabrera, B. *Phys. Rev. Lett.* **48**, 1378–1381 (1982).
10. Cabrera, B. *et al.* *Phys. Rev. Lett.* **51**, 1933–1936 (1983).
11. Incandela, J. *et al.* *Phys. Rev. D* **34**, 2637–2647 (1986).
12. Somalwar, S., Frisch, H. & Incandela, J. *Phys. Rev. D* **37**, 2403–2417 (1988).
13. Groom, D.G. *Phys. Rep.* **40**, 323–373 (1986).

tary. The problem is worse when one considers the many 'virtual' G proteins: still-unidentified mediators of distinctive patterns of cytoplasmic signalling. Perhaps a perceptive student of the cell biology of CMV infection will be able to interpret the correct signals when these CMV genes are transfected into the appropriate host cells. This student should keep his mind open to the possibility that the proteins may have some totally unexpected function that has nothing to do with G proteins.

It is fascinating to speculate on how the CMV-encoded receptors could participate in the viral life cycle. The retention of key conserved amino acids suggests that they are involved in signalling to G proteins. Do they alter host-cell metabolism to facilitate the propagation of the virus? Do they suppress specific anti-viral functions of the host cell, or apparently innocuous house-keeping activities that might detract from the production of new virions? If these receptors still respond to extracellular regulatory ligands — hormones, neurotransmitters or other regulators — they might be important in determining which cells or tissues are susceptible to

viral infection. Alternatively, these receptors might be critical in determining the productive, latent or lytic nature of the viral infection. Identifying the point in the viral life cycle when these receptors are expressed may suggest answers to these questions. Understanding their function may also tell us why a virus-infected cell is sick: a moderate production of virions by an otherwise healthy cell should not be harmful, but aberrant signals from virus-encoded receptors might contribute to the pathology of virus infection. Lastly, as Chee *et al.* point out, understanding the biology of these receptors and identifying their ligands may open an entirely new and safer road for the chemotherapy of viral infection. □

Elliott M. Ross is in the Department of Pharmacology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9041, USA.

1. Ross, E.M. *Neuron* **3**, 141–152 (1989).
2. Chee, M.S., Satchwell, S.C., Preddie, E., Weston, K.M. & Barrell, B.G. *Nature* **344**, 774–777 (1990).
3. Chee, M.S. *et al.* *Curr. Topics Microbiol. Immun.* (in the press).
4. Weston, K. & Barrell, B.G. *J. molec. Biol.* **192**, 177–208 (1986).

PROTON STRUCTURE

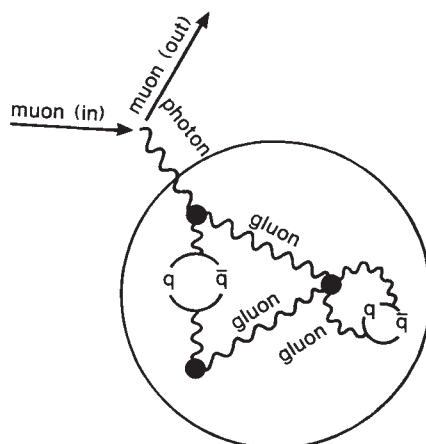
Quark spins open to question

Richard G. Roberts

It is now two years since the European Muon Collaboration at CERN (the European Particle Physics Laboratory) threw particle theorists into a spin with its measurement of the distribution of intrinsic angular momentum in the proton, which implied that the constituent quarks carried only a small fraction of the total spin¹. Even now, the debate continues heatedly with the launching of several controversial claims and counterclaims. The validity of one of the foundations of particle physics, quantum chromodynamics, has even been called into question. With so much interest focused on the structure of the proton, future guidance will probably come from two experiments currently being proposed (the Spin Muon Collaboration and the HERMES Collaboration) which will probe the spin content of both protons and neutrons.

The structure of protons and neutrons — the two types of nucleons — can be examined by observing electrons or muons scattered off them with a large transfer of momentum to one of their constituent particles or 'partons'. Such 'deep inelastic scattering' experiments provided the first confirmation of the basic structure of the proton — three valence quarks together with a sea of quarks and gluons (carriers of the nuclear strong force) (see figure). The nature of the partons is probed by the exchanged

photon. According to quantum chromodynamics (QCD), the partons interact with a force that becomes progressively weaker as the relevant momenta become larger. For this reason, scattering with a large transfer of momentum is a sensible



Probing the structure of a nucleon. A high-energy muon strikes a constituent particle in the nucleon and glances off sideways, transferring large momentum to the parton. Experimenters record the relative intensity of scattering for different values of transferred momentum and for different polarizations. It is generally agreed that nucleons comprise three valence quarks (●) together with a sea of quarks, antiquarks and gluons, and the latter can turn into quark–antiquark pairs.

way to proceed as the coupling strengths probed are sufficiently small for perturbation calculations to be used in the analysis.

According to the simple-minded view of the proton, its properties will depend on just the three valence quarks, two 'up' and one 'down', each of which carries a spin of 1/2. These will be polarized so that the up quarks contribute 4/3 of the proton's total angular momentum (which is also 1/2), and the down quark –1/3. The European Muon Collaboration (EMC) experiment² suggested that, to the contrary, all the quarks together contribute less than a third of the proton's total spin, perhaps just a few per cent. Could the remainder be carried by gluons?

The polarized experiments measure a single quantity, the structure function g_1 (a function of each parton's linear momentum, x) which is a sum of contributions from the different quark flavours:

$$g_1^p(x) = \frac{1}{2} \left[\frac{4}{9} \Delta u(x) + \frac{1}{9} \Delta d(x) + \frac{1}{9} \Delta s(x) \right]$$

(The different weightings for the three flavours, up (u), down (d) and strange (s) arise from their differing electric charges.) Determining the contributions $\Delta g(x)$ from each flavour to g_1 , and hence also to the proton's total spin, requires two further quantities (there are three unknowns). The first quantity is furnished by the inviolable 'Bjorken sum rule' which relates the proton and neutron g_1 values to quantities involved in the β -decay of the neutron (which converts a down to an up quark). The second is provided by similar values derived from the β -decay of hyperons, nucleons that contain strange (s) quarks.

The values that the EMC derived for Δu , Δd and Δs (after integrating over the linear momentum x) are 0.782, –0.472 and –0.190, respectively⁴. The sum, 12 per cent, is much less than the 100 per cent expected from the naive model, and the individual values are very different from the naive expectations, 4/3, –1/3 and 0, respectively.

Consideration of the gluons crucially changes the situation^{3–5} although the details are still debated. An important property of the gluons is their ability to turn temporarily into a 'virtual' quark–antiquark pair (see figure). If the gluons are also polarized with a distribution Δg , this modifies the apparent polarization of each of the quark flavours through the polarized virtual pairs. This modification (the so-called anomaly term) depends on the momentum transferred in the collision in the same way as the direct quark distribution itself, and so persists to high-energy collisions. The polarized gluons then make a contribution $\alpha_s/2\pi \int \Delta g(x) dx$ to the apparent polarization of each of the quark flavours, where α_s is the QCD coupling constant.

This notion has survived vigorous theoretical criticism and it also has testable experimental consequences: the quark-antiquark pairs arising from a polarized gluon can sometimes emerge from the proton to make a jet of particles with a relatively large transverse momentum. It should be possible to observe this. If the gluons are sufficiently polarized for the anomaly term to be 0.19, so that the whole of the apparent strange-quark polarization arises from the gluon distribution, the quarks would have polarizations 0.98, -0.28 and 0.0 for the up, down and strange flavours, respectively; and 70 per cent of the proton's spin would be associated with the up and down quarks. This argument can become more subtle still⁶ if one considers the linear-momentum dependence of the distributions $\Delta q(x)$, not simply the integrals.

The polarized distribution of each flavour of quark $\Delta q(x)$ is the difference between the quarks with their spins aligned with their motion and those with their spins aligned against it (that is, the difference between those with opposed 'helicities'). The distribution of polarized quarks can never exceed the distribution of unpolarized quarks, and this argument should apply to the valence and sea quarks separately. Preparata and Soffer⁷ have pushed this idea of separating distributions further, suggesting that the unpolarized sea-quark distribution itself has two components. From this they claim to show that the polarized strange-quark distribution in fact exceeds the relevant component of the unpolarized strange-quark distribution, and argue that this implies a breakdown of perturbative QCD, the usual framework for discussion of these experiments. Ellis and Karliner⁸ and, now, Close⁹ show that this line of reasoning is seriously flawed: the parting of the sea is quite arbitrary; and the calculation relies on a superseded value of one of the supplementary quantities (that relating to hyperon β -decays).

A souped-up version of the European Muon collaboration, the Spin Muon Collaboration, has been proposed, which will repeat the previous experiments but with greater accuracy and extending them to lower x . Further, it will measure the structure for the neutron as well as the proton, so testing the sacrosanct Bjorken sum rule, central to the interpretation of

the present results. And at HERA, the future West German electron-proton collider, the HERMES collaboration is proposing to scatter polarized electron beams off polarized gas targets of hydrogen, deuterium and ³He. Because the targets can be polarized both longitudinally and transversely, not only can the structure function g_1 be measured, but also an additional structure function g_2 , giving yet more detail on the proton and neutron structures.

So we may be in for more surprises. But
IMMUNOLOGY

Some savage cuts in defence

Peter Parham

TARGETED gene disruption is a fashionable tool in the molecular biologist's gadget bag; it is also a sharp one. Long familiar to yeast geneticists, the 'knockout experiment' now offers mammalian biologists the challenge of building a mouse that lacks their favourite gene¹. For 'house-keeping genes', performing basic cellular functions, it might not be worth the effort, as the probable outcome — no mouse — gives little to publish or play with. On the other hand, for genes with less central roles it can provide unique experimental systems for studying gene function. For immunologists this is wonderful news, because immune system-specific genes are involved not in keeping house, but in putting it in order when things go wrong. On page 742 of this issue Jaenisch, Raulet, Simister and their colleagues (Zijlstra *et al.*²) describe the first experiment to disrupt a gene of the immune system; and their result, that $\beta 2$ -microglobulin is not required to build a mouse, sets a precedent which will no doubt be followed by the breeding of mice lacking the other genes of antigen recognition.

$\beta 2$ -microglobulin ($\beta 2$ -m) is a 99-residue polypeptide that provides one of the two immunoglobulin-like domains of class I major histocompatibility complex (MHC) molecules³. It was the founding non-antibody member of the immunoglobulin superfamily⁴, and since its discovery has repeatedly served as a convenient molecular handle for manipulation of class I MHC molecules. For although $\beta 2$ -m itself is encoded by a single copy gene, it associates with several MHC-encoded class I heavy chains and also with other proteins, such as the CD1 thymocyte glycoproteins⁵ and the intestinal IgG receptor⁶, not encoded by the MHC. Thus by disruption of the $\beta 2$ -m gene, Zijlstra *et al.* eliminate expression and function of the entire array of class I molecules.

The first stage in making the $\beta 2$ -m negative mouse was disruption of the $\beta 2$ -m gene in a mouse embryonal stem

for the time being, it seems that QCD provides a reasonable picture of the proton's spin structure: the total spin of 1/2 could be made up of roughly 1/3 from the quark-partons and a remainder which may be regarded as an approximate cancellation between two large quantities — the gluon helicity contribution and a (negative) orbital piece. □

Richard G. Roberts is in the Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK.

The immune system as a defence organization

1. Its function is selective destruction.
2. It is large, complicated and elaborate.
3. It is expensive.
4. It is wasteful.
5. It has distinct components performing apparently identical functions.
6. It is slow to react.
7. It is prepared for events that never happen.
8. It fights today's threats with the solutions of past problems.
9. It is susceptible to corruption.
10. It can destroy that which it protects.

cell line, through homologous recombination with an engineered, non-functional $\beta 2$ -m gene (see also Capecchi in News and Views¹). These cells were then incorporated into mouse embryos, which developed to give germ-line transmission of the disrupted gene. Mice heterozygous for the disrupted gene are phenotypically indistinguishable from wild-type animals and in themselves are of little interest. Yet such is the novelty — and the promise — of mammalian gene disruption that Zijlstra *et al.*, and their competitors Koller and Smithies, paused in mid-experiment to report on the heterozygotes^{7,8}.

Science takes over from technology with breeding of the heterozygotes. And in the case of $\beta 2$ -m Zijlstra *et al.* obtained viable homozygotes: "apparently healthy mice" that "were indistinguishable from heterozygous or wild type littermates". But unlike their littermates the homozygotes are devoid of $\beta 2$ -m and have almost completely lost cell-surface expression of class I heavy chains. The integrity of these mice provides the final nail (there are many others, as recently discussed in News and Views⁹) in the coffin of claims that cell-cell interactions in embryogenesis are programmed by a sequence of $\beta 2$ -m-associated MHC-like molecules¹⁰. Although such schemes have long been discredited, the report by Zijlstra *et al.* marks once and for all the end of

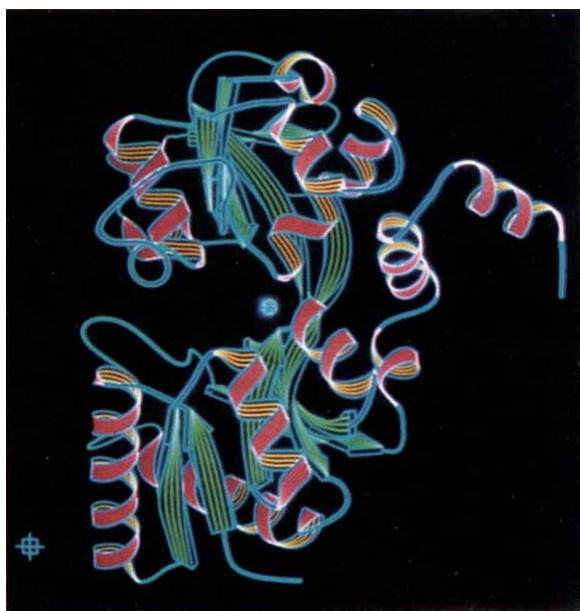
1. Sachrajda, C. T. *Nature* **333**, 703–704 (1988).
2. Ashman, J. *et al. Nucl. Phys.* **B328**, 1–35 (1989).
3. Altarelli, G. & Ross, G. G. *Phys. Lett.* **B212**, 391–395 (1988).
4. Carlitz, R. D., Collins, J. C. & Mueller, A. H. *Phys. Lett.* **B214**, 229–236 (1988).
5. Efremov, A. V. & Teryaev, O. V. *Dubna preprint JINR-E2-88-287* (1988).
6. Altarelli, G. & Stirling, W. J. *Part. World* **1**, 40–46 (1989).
7. Preparata, G. & Soffer, J. *Phys. Rev. Lett.* **61**, 1167–1169 (1988).
8. Ellis, J. & Karliner, M. *Phys. Lett.* **B213**, 73–80 (1988).
9. Close, F. E. *Phys. Rev. Lett.* **64**, 361–365 (1990).

The jaws of lactoferrin

TRAPS come in all shapes and sizes, but most work on the 'lure and snare' principle. The iron-binding protein lactoferrin, one of whose two jaw-like 'iron traps' is depicted in the colour figure, is no exception. On page 784 of this issue, Edward Baker and colleagues of the Massey University in New Zealand show that in the absence of any bound metal one of lactoferrin's two traps lies wide open, with its molecular jaws poised to close firmly around an incoming iron cation.

Lactoferrin, like all members of the transferrin family of proteins, plays an essential part in regulating levels of vagrant, free iron inside animals. Earlier studies of the molecular structure of lactoferrin bound to iron showed that it consists of two similar, adjoining protein lobes each of which encloses a single trapped iron cation. But whether these lobes are permanently closed or whether their closure is triggered by iron binding was left unanswered.

By determining the crystal structure of apolactoferrin, Baker and colleagues now show that in the absence of trapped iron the hinged protein domains of one of the two lobes swing open, whereas the second lobe remains firmly closed. Their findings imply that there is only a small energy difference between the open and closed states, and that minor perturbations of lactoferrin's structure such as those resulting from crystal packing forces or the binding of iron are sufficient to trigger the mechanical closure of the protein's molecular jaws. Lactoferrin, it seems, can justifiably be likened to a Venus fly trap.



David Concar

a formerly influential concept, a clarification of the collective wisdom and a lightening of the literature.

Although many cellular and molecular phenomena have been associated with $\beta 2$ -m and $\beta 2$ -m-associated class I molecules, there are only two substantiated functions and they are both to do with immunity. One is the presentation of antigens by classical class I MHC molecules (H-2K, D and L in mice), to T lymphocytes bearing $\alpha\beta$ antigen receptors and the CD8 coreceptor; the other is the trans-epithelial transport in neonates of IgG from mothers' milk by the $\beta 2$ -m-associated Fc receptor. In the $\beta 2$ -m negative mouse both functions have disappeared. IgG does not bind to the epithelial brush border membranes of neonatal intestines. The peripheral lymphoid organs have less than 1 per cent of the normal number of CD8⁺ $\alpha\beta$ -receptor-bearing T cells, and cytotoxic functions that are normally provided by these cells cannot be detected. How does one reconcile such damage to the immune system with the paradoxical vigour of the mice?

Specific immunity is found only in organisms having many specialized tissues and cell types. In the past, when some

level of organismal complexity was attained it clearly became advantageous to evolve cells and molecules that were increasingly committed to protection and defence. An analogous progression can be seen in the evolution of military defence as human societies have become increasingly complex and individuals more specialized and interdependent (see box, on previous page).

Two key features distinguish components of the specific immune system from those of cells and tissues that have more constructive roles. First, they are used sporadically, only when something goes wrong. Second, the opponents of the immune system — bacteria, viruses, parasites — are unpredictable and rapidly changing, so that past experience does not necessarily prepare the immune system for the future. Thus the selective forces acting upon specific immune functions are constantly varying and this leads to products that are often poorly adapted to a particular set of circumstances. Hence the continuing loss of human life from infectious diseases, the incapacitation resulting from the all too common cold and our frightening vulnerability to HIV.

The fitness in the face of immunodeficiency

of Zijlstra *et al.*'s mice confirms the view that the immune system has an essentially defensive role — that it is only needed when things go wrong. Naturally, the authors have been careful to protect their first homozygotes from attack by microorganisms; the animals are kept in a pathogen-free environment, and are not required to forage. But as numbers of mice accumulate, Jaenisch and colleagues will be able to change their tactics and challenge the immune system of their $\beta 2$ -m negative mice. Cytotoxic CD8⁺ cells kill virally infected cells and are involved in limiting virus spread and thereby ending infections. In contrast it is soluble antibody that often prevents initiation of an infection. One might therefore expect that $\beta 2$ -m negative mice will show poor recovery from viral infections, but on specific immunization their B cells and CD4⁺ helper T cells will provide normal protection from viral infection¹¹.

Using transgene technology, immunologists have been able to introduce exogenous genes into mice and see how they perturb the immune system; with targeted gene disruption they can now take out the endogenous genes. Combination of these two techniques will prove geometrically powerful and lead to an even more detailed understanding of the workings of the immune system. For example the function(s) of $\beta 2$ -m can now be finely dissected by making the $\beta 2$ -m negative mice transgenic for $\beta 2$ -m genes containing engineered mutations.

Studies with transgenic mice have demonstrated positive thymic selection of $\alpha\beta$ -receptor-bearing T cells by class I and class II MHC molecules¹². The $\beta 2$ -m negative mice provide a particularly clear illustration of this previously controversial mechanism: without class I MHC molecules there is no selection and consequent migration into the peripheral lymphoid tissues of CD8⁺ T cells with $\alpha\beta$ receptors. In contrast there is no effect on the numbers of peripheral $\gamma\delta$ -receptor-bearing T cells, including those expressing CD8. T cells with $\gamma\delta$ receptors are much less frequent than their companions with

1. Capecchi, M. *Nature* **344**, 105 (1990).
2. Zijlstra, M. *et al.* *Nature* **344**, 742–746 (1990).
3. Möller, G. *Transplantation Reviews* (Munksgaard, Copenhagen, 1974).
4. Peterson, P.A., Cunningham, B.A., Berggard, I. & Edelman, G.M. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1697–1701 (1972).
5. Calabi, F. & Milstein, C. *Nature* **323**, 540–543 (1986).
6. Simister, N.E. & Mostov, K.E. *Nature* **337**, 184–187 (1989).
7. Zijlstra, M., Li, E., Sajadi, F., Subramani, S. & Jaenisch, R. *Nature* **324**, 435 (1989).
8. Köller, B.H. & Smithies, O. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8932–8935 (1989).
9. Goodfellow, P.N. *Nature* **343**, 597–598 (1989).
10. Bennett, D. *Cell* **6**, 441–454 (1975).
11. Mims, C.A. *The Pathogenesis of Infectious Disease* (Academic, London, 1987).
12. Von Boehmer, H., Teh, H.S. & Kisielow, P. *Immunol. Today* **10**, 57–61 (1989).
13. Strominger, J.L. *Cell* **57**, 895–898 (1989).
14. Lawlor, D.A., Zemmour, J., Ennis, P.D. & Parham, P. *Ann. Rev. Immunol.* **8**, 23–63 (1990).

$\alpha\beta$ receptors, and their function is poorly understood. As Zijlstra *et al.* delicately intimate, this calls into question hypotheses that $\gamma\delta$ T-cell receptors are thymically selected to interact with CD1 and various non-classical class I MHC antigens, all of which are $\beta 2$ -m associated¹³. As cellular immunologists acknowledge, almost any desired T-cell specificity can be found providing one looks hard enough. So the observed interactions between $\gamma\delta$ T-cell receptors and non-classical class I molecules might be fortuitous

cross-reactions, such as are occasionally found with monoclonal antibodies. The evidence now accumulating suggests that many non-classical class I molecules are evolutionary remnants of little functional consequence for the contemporary immune system¹⁴. The same may well be true for the $\gamma\delta$ T-cell receptor: time and targeted gene disruption will tell. $\square$

Peter Parham is in the Department of Cell Biology, Stanford University, Stanford, California 94305, USA.

GREENHOUSE EFFECT

Satellite data under scrutiny

P. D. Jones and T. M. L. Wigley

A NEW study, published in *Science*¹, presents the results of ten years of global temperature estimates from satellite-based radiometers. In their analysis, Spencer and Christy say that their estimates show no evidence of warming over this ten-year period. This conclusion, reasonable enough in itself, is unfortunate because taken out of context it is open to the interpretation that the authors have found no evidence for global warming in general — that indeed is the way the study has been widely reported. Yet the data themselves are of considerable value and add another dimension to the analysis of temperature trends.

The radiometers (microwave sounding units or MSUs) were installed on the TIROS-N series of satellites first launched by the National Oceanic and Atmospheric Administration (NOAA) in late 1978. They measure the temperature-dependent thermal emission of oxygen at four frequencies near 60 gigahertz; as oxygen concentrations are relatively constant in space and in time, the emissions can be interpreted as indicators of temperatures. The different frequency bands or channels can be combined to give an indication of the average temperature of the lower atmosphere from the surface to around

10 km. These data can then be spatially averaged to estimate global-mean values. Cross calibration of the results of the NOAA-6 and NOAA-7 satellites over the period when they were both in orbit enables the length of the temperature record to be extended to ten years.

In assessing these data, one must be careful to avoid comparing apples and oranges. The MSU data represent an unequally weighted average temperature of the troposphere, with the largest weight on the zone of the atmosphere which lies between the pressure levels of 850 and 500 hectopascals, about 1.5 to 5 km. It would therefore be most appropriate to compare these data with conventional radiosonde data for the free atmosphere. Instead, Spencer and Christy make comparisons with surface data. This is still useful — after all apples and oranges are both round fruits, and surface and free atmosphere data do correlate well² — but the comparison fails to do justice either to the satellite data or the surface data. When global annual mean MSU data are correlated with mean temperatures between 850 and 300 hPa (ref. 3), the agreement is excellent ($r^2=0.94$, $n=10$), providing mutual support for both types of data (see table). The correlations become less

strong on smaller spatial and temporal scales — understandably so, because the MSU data do not fully reflect the 850–300 hPa mean temperature.

Correlations with the land-based^{4,5} and land-plus-marine temperature data⁶ are, as one would expect, less good. But they are better than the values given by Spencer and Christy, whose correlations are not always based on the full ten-year period. For global annual mean data the r^2 value is 0.86 (see table). The satel-

lite data therefore show that worries about incomplete spatial coverage or data distortion due to urban warming effects in the ground-based data are probably unfounded. Further comparisons are required to assess the full value of the satellite data as an indicator of what is happening near the Earth's surface, and, in turn, to address the question of the coverage and quality of the surface data more fully.

The authors claim that the radiometers should provide the standard for monitoring global temperatures. It would be more prudent, however, to consider them as complementary to the ground-based data, not as a replacement. Spencer and Christy also claim that differences between the various research groups^{7,8} who have calculated hemispheric-mean temperature values have led to conflicting reports in the popular press.

This may be true, but the data sets themselves are not in conflict. The two land-data sets^{4,5} are in excellent agreement ($r^2 = 0.93$ for global annual means over the 1881–1980 period). Confusion has arisen, however, with regard to the use, or otherwise, of marine data. The Hansen–Lebedeff⁸ data are land only whereas the Jones *et al.* data^{6,7} include both land and marine observations. Agreement between land-only and land-plus-marine data is also good^{2,8}, which is a further indication that incomplete coverage is not a serious problem.

The conclusion that there is no evidence of a warming trend over the past ten years contributes little to the debate about global warming in the long-term, in which it is the time series of temperature over the past century that must be considered. The land, marine and upper-air temperature records (see table) agree with the satellite data in showing no statistically significant increase in temperature over the past ten years. But these longer records do show that, relative to the earlier part of the century, the 1980s were unusually warm, being the warmest decade on record and 0.2 °C above the 1950–79 average. $\square$

P. D. Jones and T. M. L. Wigley are in the Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK.

Monthly, seasonal and annual explained variances between the MSU satellite temperatures and surface and 850 and 300 hPa temperature estimates, all calculated over 1979–88.

	NH (%)	SH (%)	GL (%)
Monthly			
Land ^{4,5}	23	11	32
Land and marine ⁶	25	20	38
Seasonal			
Land ^{4,5}	37	32	47
Land and marine ⁶	43	41	58
850–300 ³	58	35	60
Annual			
Land ^{4,5}	68	73	74
Land and marine ⁶	70	60	86
850–300 ³	74	72	94

NH, Northern Hemisphere; SH, Southern Hemisphere; GL, global.

1. Spencer, R. W. & Christy, J. R. *Science* **247**, 1558–1562 (1990).
2. Wigley, T. M. L., Angell, J. K. & Jones, P. D. in *Detecting the Climatic Effects of Increasing Carbon Dioxide* (eds MacCracken, M. C. & Luther, F. M.) 55–90 (US Department of Energy, Washington, DC, 1985).
3. Angell, J. K. *J. Clim.* **1**, 1296–1313 (1988).
4. Jones, P. D. *et al. J. Clim. appl. Meteor.* **25**, 161–179 (1986).
5. Jones, P. D., Raper, S. C. B. & Wigley, T. M. L. *J. Clim. appl. Meteor.* **25**, 1213–1230 (1986).
6. Farmer, G., Wigley, T. M. L., Jones, P. D. & Salmon, M. *Detecting and Explaining Recent Global Mean Temperature Changes* (Climatic Research Unit, University of East Anglia, Norwich, 1989).
7. Jones, P. D., Wigley, T. M. L. & Wright, P. B. *Nature* **322**, 430–434 (1986).
8. Hansen, J. & Lebedeff, S. *J. geophys. Res.* **92**, 13345–13372 (1987).

Rhizobium sweet-talking

Sharon R. Long and E. Morrey Atkinson

ONE of the most dramatic changes of developmental fate seen in any plant system occurs when the roots of leguminous plants abruptly alter their normal form of cell division, growth and differentiation to create organs called nodules¹. Nodule development in these plants is caused by specific bacteria in the *Rhizobium* group, which invade the emerging nodule and establish a nitrogen-fixing symbiosis with the plant. Researchers have energetically sought to find out how *Rhizobium* causes nodulation, because the answer may provide clues about fundamental processes in plant development and cell recognition. As reported by Lerouge *et al.* on page 781 of this issue², researchers at two CNRS laboratories in Toulouse have now determined the structure of a *Rhizobium*-synthesized molecule, NodRm-1, which could be the signal that flips the developmental switch in the plant host.

The structure of NodRm-1 is a novel one: it is a β -linked glucosamine tetrasaccharide, with three of the sugars bearing an *N*-acetyl group; the unit on the reducing end is modified by a sulphate, and the unit at the non-reducing end bears an *N*-acyl C-16 fatty acid modification. This molecule is therefore unlike any of the main endogenous plant hormone types, such as cytokinins, gibberellins or auxins. The finding that NodRm-1 is a modified oligosaccharide is of great interest, given current research on oligosaccharides as plant intercellular signals³ and on involvement of host lectins in nodulation⁴.

The initial stages of interaction of *Rhizobium* and plants include the deformation of growth in plant epidermal root hairs, and the initiation of cell division in the cortex of the host root (Fig. 1a and

b). These steps occur only if the *Rhizobium* is of the correct host-range type. The organism studied by Lerouge *et al.* is *R. meliloti*, which invades alfalfa, but not vetch or pea. Genetic analyses of numerous *Rhizobium* species have revealed loci that are required for nodulation, termed *nod* genes. These genes fall into two general groups. Common *nod* genes include *nodABC*, and seem to be required in all *Rhizobium* systems for eliciting root-hair deformation and root-cell division (Fig. 1b). Both of these types of plant reaction may be caused by signals whose production requires *nodABC*; alternatively, one reaction may be caused by *nodABC*-dependent signals, with the other being a secondary effect^{5,6}. A second group of *nod* genes influence *Rhizobium*'s host range, that is, which particular plants can be nodulated^{7,8}. Such genes include *nodH*, which is required for *R. meliloti* to infect its own host plant, alfalfa, but is not found in other *Rhizobium* species such as *R. leguminosarum* biovar. *viciae*, the pea and vetch plant symbiont.

The idea that *Rhizobium* produces extracellular factors that control plant morphogenesis dates back many years, and was supported as early as 1969 by the finding that cell-free filtrates from *Rhizobium* cultures could induce root-hair deformation in a way that parallels the host range of the live cells⁹. More recent studies have been based on use of *Rhizobium nod* gene mutants and strains with cloned *nod* genes, and on the discovery that *nod* genes are regulated by plant

signals, known to be various flavonoids¹. For example, Van Brussel *et al.*¹⁰ established that *R. leguminosarum* biovar. *viciae* filtrates had an effect on the root morphology of vetch. They showed that the activity of these culture filtrates depends upon the presence of *nod* genes, and upon pre-exposure of the *Rhizobium* cells to plant flavonoid inducers.

A similar line of research has been followed in Toulouse, where the effect on plants of *R. meliloti* strains carrying various *nod* genes was found to parallel the activities of cell-free filtrates from those strains. The general model of Faucher *et al.*¹¹ is shown in Fig. 2: *R. meliloti* carrying all of its *nod* genes produces extracellular factor(s) that deform growing root hairs on alfalfa; if *nodABC* are mutated,

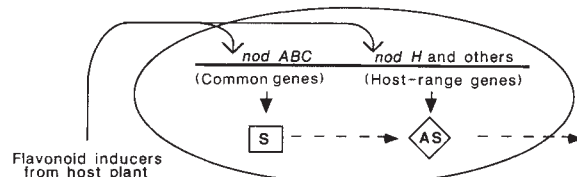


FIG. 2 The model developed by Faucher *et al.*¹¹ through bioassay of cell-free filtrates of *Rhizobium meliloti*. The common *nod* genes, *nodABC*, are responsible for production of a basic, common signal (S), which causes hair deformation on a non-host plant, vetch. Through the action of the host range genes, which include *nodH* and *nodQ* (ref. 11), this common signal is converted to an alfalfa-specific signal (AS), which deforms hairs on alfalfa but not on vetch.

then the live *Rhizobium* cells cannot deform host root hairs, nor can the cell supernatants cause root-hair deformation. (Had — hair deformation — activity is a commonly used phenotype for *Rhizobium* symbiotic studies.) But mutations in *nodH* do not completely abolish production of extracellular factor. The specific activity of the culture filtrate on alfalfa has been lost, but a non-specific factor with Had activity on vetch plants is still present. The Had bioassay was followed as a means to purify the alfalfa-specific *nod* gene factor. A series of spectroscopic and other analyses revealed the modified oligosaccharide structure. Future experiments may need a broader approach to the chemistry: the process of fractionation and bioassay that here made the final identification of NodRm-1 possible will by definition not reveal inactive intermediates, nor will it detect products of other *nod* genes that have no intrinsic effect on root-hair growth.

How is NodRm-1 synthesized? Knowledge of the structure will be a tremendous aid to pursuit of the synthetic pathway. An interesting lead pointed out by Lerouge *et al.* is the similarity of the NodRm-1 structure to the structures of intermediates in bacterial lipopolysaccharide and cell-wall synthesis. Another aspect of the pathway question is whether the β -glucan backbone might be constructed by a polymerase activity, and whether this polymerase stops specifically

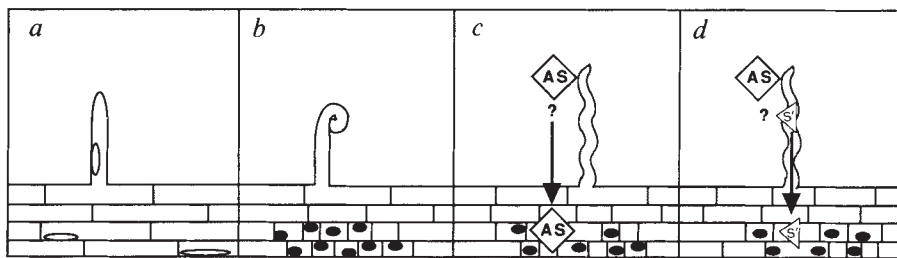


FIG. 1 Representation of plant reactions to *Rhizobium* and models for signalling, modified from ref. 6. a, Uninfected roots show straight root hairs and large vacuolate cortical cells. b, In alfalfa plants infected by *Rhizobium*, growth of epidermal root hairs is altered into a curled form, suggesting drastic effects of the bacteria on the deposition of the extracellular matrix of the root hairs. In addition, root cortical cells begin dividing, several cell layers distant from the bacteria on the surface. However, plants inoculated with *Rhizobium* carrying *nodABC* mutations show neither effect, and resemble the plant shown in (a). c, The alfalfa-specific signal (AS), termed NodRm-1 by Lerouge *et al.*, causes root hair deformation on alfalfa. Does it stimulate plant cell division? One possible mechanism could be a direct interaction of the signal with cells of the inner cortex. d, Another means by which the NodRm-1 signal or other alfalfa-specific signals could cause plant cell division would be by transduction of the original bacterial signal (AS), and release of an endogenous secondary signal, S', which causes initiation of cell division deeper in the root.

after four residues, or whether it constructs polymers of higher molecular weight which are subsequently cleaved into the active tetrasaccharide units. Ultimately, the synthetic pathway will have to be reconstructed *in vitro*, using the protein products of the *nod* genes already cloned and sequenced.

Are all of the known *nod* genes somehow involved in NodRm-1 synthesis, or do some of them carry out other functions? Knowledge of the NodRm-1 structure makes it possible to address these questions directly. A bonus will be the clues provided for researchers studying other *Rhizobium* species, who can now search for the synthesis of structurally related signal molecules in their systems. Because the *nodABC* genes are universally involved in *Rhizobium* nodulation, signal molecules in each *Rhizobium* probably have some basic features in common, through enzymatic activities provided by *nodABC* gene products.

And how exactly does the plant respond to this signal? NodRm-1 was identified by assay of root-hair deformation; the idea that *nod*-gene factors also affect cell division is mostly an inference from analysis of mutant phenotypes^{1,5}. Does NodRm-1 cause cell division in the roots of the plant host? If so, this might be a direct effect, or it might be caused through transduction of the NodRm-1 information into a secondary signal (Fig. 1, compare *c* and *d*). Among the candidates for secondary signals would be endogenous plant hormones — nodule-like structures can form in response to anti-auxins¹² and to bacteria producing cytokinins⁶, so an alteration in auxin-cytokinin ratio in roots could be one mechanism for nodule induction. But it is also possible that *Rhizobium* signals cause cell division in plants through fundamental plant cell mechanisms that are as yet unknown, but which are mimicked by anti-auxins or cytokinins. Among the advances needed for understanding the action of NodRm-1 will be the description of other cellular effects of this signal on plant cells, and the characterization of the plant receptor.

Until the NodRm-1 synthetic pathway is established, the actions of various *nod*

genes are accounted for, and the receptor (and internal effectors?) in the plant are found, the sufficiency of NodRm-1 for alfalfa nodulation will be uncertain. Even so, the determination of a specific structure for a nodulation signal has long been a rate-limiting step for nodulation research, and the result reported here will open the door on a new era of experiments. NodRm-1 will be a useful tool for ASTROPHYSICS

investigating the inner workings of plant cells, and for asking how these cell mechanisms are subverted to accomplish the transition from root growth to nodule development. □

Sharon R. Long and E. Morrey Atkinson are in the Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA.

Black holes in the balance

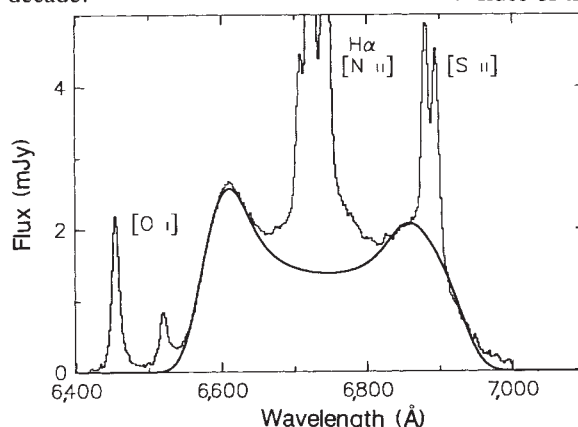
Jules P. Halpern

FOR more than 25 years, astrophysicists have clung to the belief that the accretion of mass onto black holes, and the accompanying conversion of gravitational energy into radiation, are responsible for some of the most energetic displays in the Universe. Two astronomical settings thought to harbour black holes are active galactic nuclei (which include quasars, radio galaxies and Seyfert galaxies), and a small fraction of the binary stellar X-ray sources in our own Galaxy, which for a variety of reasons are called black-hole 'candidates'. But science advances neither by faith nor by votes. The demonstration of the existence of a black hole requires evidence of its compactness, ideally the observation of relativistic effects in the strong gravitational field, and a measurement of its mass which has to be large enough to exclude other compact objects such as neutron stars (theoretically, neutron stars are unlikely to exceed 3.6 solar masses). On page 747 of this issue Stella¹ outlines a practical technique whereby such a proof may be achieved using new observational data that will be acquired in the coming decade.

Our last glimpse of material falling into a black hole is probably obtained during its inward spiral through the accretion disk, a circular plane of hot gas, differentially rotating in nearly keplerian (planetary) orbits, and the source of a large fraction of the radiation which we see from active galactic nuclei (AGNs) and binary X-ray sources. The search for black holes has been intimately tied to the study of accretion disks, and sufficient understanding of the properties of these disks could be used to infer the existence of black holes. The recent developments inspiring hope for a precision measurement of the mass of a black hole involve the detection of spectral lines, both in AGNs and binary X-ray sources, which appear to arise on the surface of an accretion disk.

The figure shows a very broad Balmer α line of hydrogen, on which is superimposed a double-peaked model profile characteristic of a disk-like, rotating configuration of emitting material. The two broad horns are due to Doppler shifts in the wavelengths observed from the material on the approaching and receding sides of the disk. The asymmetries in the peaks and the wings are due to relativistic effects. This particular observation is of visible radiation from an AGN. In X-ray binaries, the broad lines which have been seen are due to K α emission from iron at X-ray wavelengths. But the spectral resolution of the present X-ray data is not quite adequate to resolve the double peaks (if they exist). X-ray lines have been discovered in several AGNs as well, although their widths have not yet been measured.

The usefulness of these data lies in the information they carry about the orbital velocities, the gravitational redshift and the bending of light rays in the strong gravitational field of the black



The broad emission line profile of the AGN Arp 102B. The bold line represents a relativistic disk model with inner and outer radii of 175 and 500 times the Schwarzschild radius (event horizon) respectively. The asymmetries in the profile are due to relativistic effects. The narrow lines in this spectrum come from elsewhere in the system, and can be ignored for the purpose of the accretion disk model. Time variations in a line profile such as this could be used to measure the mass of the putative black hole.

- Long, S.R. *Cell* **56**, 203–224 (1989).
- Lerouge, P. et al. *Nature* **344**, 781–784 (1990).
- Eberhard, S. et al. *Plant Cell* **1**, 747–755 (1989).
- Diaz, C.L., Melchers, L.S., Hooykaas, P.J.J., Lugtenberg, B.J.J. & Kijne, J.W. *Nature* **338**, 579–581 (1989).
- Dudley, M. & Long, S.R. *Plant Cell* **1**, 65–72 (1989).
- Long, S.R. & Cooper, J. in *Molecular Genetics of Plant Microbe Interactions* (eds Palacios, R. & Verma, D.P.S.) 163–178 (APS Press, St Paul, Minnesota 1988).
- Debellé, F. et al. *J. Bact.* **168**, 1075–1086 (1986).
- Horvath, B. *Cell* **46**, 335–343 (1986).
- Yao, P.Y. & Vincent, J.M. *Plant Soil* **45**, 1–16 (1969).
- Van Brussel, A.A.N. et al. *J. Bact.* **165**, 517–522 (1986).
- Faucher, C., Camut, S., Denarie, J. & Truchet, G. *Molec. Plant-Micro Interact.* **2**, 291–300 (1989).
- Hirsch, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1244–1248 (1989).

hole. K. Chen² has described how the observed line profiles can be used to infer various properties of the accretion disk. If the disk were perfectly smooth and unvarying, it would not be the mass of the central black hole that would be determined, but the ratio M/R , where M is the central mass and R is the radius of the disk. But nature is not so perfect (fortunately, in this case), and ubiquitous fluctuations in luminosity could, in principle, provide the extra piece of information needed to determine both M and R . In several AGNs it is known that variations in the luminosity of the nucleus are echoed, several days later, in the strengths of the emission lines. Imagine, as Stella suggests, that a source of ionizing radiation at the centre of an accretion disk illuminates the surface in bursts. With each burst, a wave of fluorescent emission will travel over the surface at the speed of light, disrupting the observed line profile in an organized and predictable way. Bumps will drift across the line profile as the illuminated region of the disk moves from the inner, high-velocity boundary to the outer, low-velocity edge. The drift time reveals the absolute radius of the disk, and hence the mass of the black hole. Stella's proposal is an application of the reverberation mapping technique, first elaborated by Blandford and McKee³, which he generalizes to include relativistic effects.

In reality, a number of doubts cloud this idealistic vision. Among the AGNs, only a handful have Balmer line profiles that bear any resemblance to the simple disk models, so it is by no means certain that the bulk of the line emission arises in disks. Even a double-peaked, asymmetric profile is not uniquely attributable to a relativistic keplerian disk. But an independent test involving spectropolarimetry of the line has been proposed^{2,4}, and could be carried out with existing instruments. In X-ray binaries, the as-yet-undetermined shape of the iron K α emission lines could be affected by non-gravitational effects, such as Compton scattering in a hot corona above the disk, and several different ionization stages of iron may contribute to an unresolved blend of closely spaced lines. On the optimistic side, the iron lines observed in Cygnus X-1 — the prototype black-hole candidate — and the X-ray transient 4U 1543-47, are too broad and too low in energy to be described as a blend of different ionization stages. Doppler and gravitational redshifts near a black hole remain the most probable explanation. Also, detailed models for iron lines that would be produced in an accretion disk⁵ seem to indicate that rotational broadening is more important than the effects of Compton scattering in a hot corona, at least for the properties of the coronae that are thought to exist in X-ray binaries.

So there is ample hope that Stella's procedure for weighing a black hole will be implemented. Black-hole masses in AGNs are expected to range between 10^6 and 10^{10} solar masses. For a 'typical' object of 10^8 solar masses, the visible line-emitting portion of the accretion disk is perhaps 1,000 times the size of the Earth's orbit. The spectrum should be monitored at daily or shorter intervals to observe the drifting features (predicted to cross the disk at the speed of light). This is no more difficult than a recent, successful international campaign to observe a Seyfert galaxy once every four days for seven months. If the X-ray lines in AGNs also come from the disk, then they will arise even closer to the black hole, requiring more frequent sampling to follow the drifting features. Although the relativistic effects are more extreme for the X-ray lines, the greater precision and light-gathering power of optical spectroscopy now allow a more accurate determination of the disk's parameters. However, the energy resolution of the order of 2 per cent which is anticipated for the next generation of X-ray satellites (such as Astro-D, AXAF and XMM), will also be adequate to test relativistic disk models. In the case of X-ray binaries, with black holes of perhaps ten solar masses, light travel times across the disk are a few milliseconds, so that it may be impossible to follow individual drifting features. Nevertheless, cross-correlation methods which compare time delays between different portions of the line profile may yield the same information. And in X-ray binaries, the same techniques would apply if the compact object were a weakly magnetized neutron star, for which the inner edge of the accretion disk could reach close to the stellar surface.

Currently, the best way to elevate an X-ray source to the exalted status of black-hole candidate is to determine a sufficiently large lower limit for its mass, using the orbital parameters of the binary. It is hard to argue against Kepler's laws. But to achieve this, some estimate of the mass of the companion star is necessary. Stella's method obviates the role of the companion star entirely, making direct use of the strong gravitational field near the black hole, and making it possible for the test to be applied to any compact object with line emission from an accretion disk. $\square$

Jules P. Halpern is in the Department of Physics, Columbia University, New York, New York 10027, USA.

1. Stella, L. *Nature* **344**, 747-749 (1990).
2. Chen, K. *Bull. Am. astr. Soc.* **22**, 757 (1990).
3. Blandford, R.D. & McKee, C.F. *Astrophys. J.* **255**, 419-439 (1982).
4. Chen, K. & Halpern, J.P. *Astrophys. J. Lett.* (in the press).
5. Kallman, T. & White, N.E. *Astrophys. J.* **341**, 955-960 (1989).

Instant metal

MANY traditional technologies, from plastering to wedding-cake decoration, employ a fine powder which rapidly sets when made into a paste with water. Some of these powders, like cement and plaster, react with the water; others, like mud and icing sugar, do not. In either case, the paste sets by the growth and interlocking of small needle-crystals throughout the drying material.

In this connection Daedalus recalls how a chemist 'ripens' a precipitate for easy filtering by keeping it suspended in excess solvent. Thanks to its high surface energy, a fine-grained precipitate is slightly more soluble than a coarser one. So fine crystals slowly dissolve, giving a solution supersaturated with respect to coarser crystals. These precipitate out, freeing the solvent to dissolve more fine crystals. Even a highly insoluble precipitate can be rapidly coarsened in this way. So, says Daedalus, almost any fine powder, wetted with a solvent that dissolves it slightly, should 'ripen' to a hard cement, provided its crystals interlace strongly as they grow.

Inspired by this reasoning, DREADCO's chemists are seeking a metallic cement. They are preparing ultra-fine metal powders by mechanical grinding, chemical precipitation and vapour deposition, and wetting them with a variety of solvents. Few solvents dissolve metals at all well. The best are probably liquid ammonia and mercury (as used in dental amalgam, an existing metallic cement). Neither is an ideal component for a safe commercial cement. But with good fortune some fairly harmless chemical relative of ammonia will show the modest solvent ability required, and DREADCO's 'Metallic Plaster' will become a commercial reality.

It should be widely welcomed. The huge electronics business will drop hot soldering with a sigh of relief, and take to cold plastering instead. Garages, plumbers and light-engineering firms will similarly lay aside their blow-lamps and welding torches. Domestic handymen will have a wonderful new material for their traditional bodging, and dentists will gladly abandon poisonous mercury amalgam for hygienic silver-plaster fillings.

Metallic Plaster may even become a material of construction in its own right. Effortlessly moulded, but almost as strong as dense metal, it will be ideal for prototype components and short production runs. Its residual porosity could be 'filled' by immersion in a lower-melting metal, giving a sort of coarse alloy, or even in a plastic monomer or ceramic slip, giving a totally novel metallic composite. Even on its own, Metallic Plaster may encourage a reflowering of the wonderful rococo engineering decorations that the Victorians used to make out of cast iron. David Jones

Feature-blind grammar and dysphasia

SIR—Developmental dysphasia — the inability of apparently normal children to acquire language normally — is well-known, but precisely what is wrong with their language and the causes of their errors are not. I now describe the outcome of testing for dysphasia the members of a

29 but that for dysphasics only 18.

While the language of dysphasic adults seems normal at first sight, careful testing shows that the normality is only apparent. Those affected may have learned strategies for coping with language, but their underlying grammar is still severely

cerned show up in spontaneous speech, writing, grammatical judgement and repetition. Because the deficits are apparent in all aspects of language, their roots probably lie in the underlying grammar rather than in a peripheral processing system. Because the language skills that are not impaired are at least as complex as those which are, it is unlikely that the underlying deficit is one of cognition as such.

The distribution of dysphasia in this one large family suggests that it may be due to one dominant gene; M. Pembry (Institute of Child Health, London) is searching for appropriate genetic markers. Though there has not, to my knowledge, previously been such a clear family history of dysphasia, there have been anecdotal observations of dysphasia in different members and different generations of the same family.

Naturally, a single family cannot prove that the dysphasia described here is genetic, but we shall soon be undertaking a study of the patterns of occurrence of dysphasia in families and testing our feature-blindness hypothesis on a larger scale. A similar study of genetic factors in severe language disorders is being followed by B. Sahlen (Lund University, Sweden); R. Frackiowak (MRC Cyclotron Unit, Hammersmith Hospital, London) and D. Bub (Montreal Neurological Institute) will be investigating the neurological correlates of the disorder and F. Vargha-Khadem (Institute of Child Health) will be looking at cognitive aspects of the deficit.

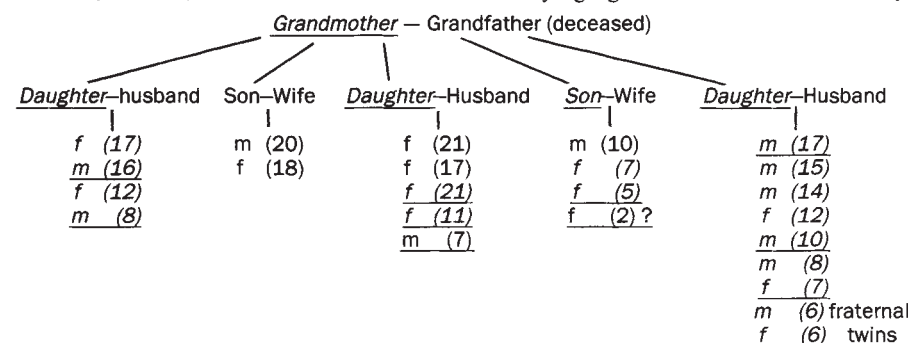
M. GOPNIK

*Department of Linguistics,
McGill University,
1001 Sherbrooke Street West,
Montreal, Quebec H3A 1G5, Canada*

1. Paradis, M. in *The Assessment of Bilingual Aphasia* (Erlbaum, Hillsdale, New Jersey, 1987).
2. Clahsen, H. *Linguistics* **27**, 897–920 (1989).
3. Gopnik, M. *A Featureless Grammar in a Dysphasic Child* (McGill Working Papers in Linguistics, McGill University, Montreal, 1986).
4. Gopnik, M. in *Language Acquisition* (Erlbaum, Hillsdale, New Jersey, in the press).

Folding proteins

SIR—Recombinant DNA technology and protein engineering make it feasible to produce substantial quantities of any protein with a specified amino acid sequence, but many potentially useful applications are impeded because the protein is produced in inclusion bodies, in an insoluble unfolded form that must be solubilized and folded to yield a biologically active product¹. Often, for reasons which are not clear, the solubilized protein will not fold as readily as would be expected. Our experience with the production of recombinant bovine pancreatic trypsin inhibitor (BPTI) may suggest a possible explanation for inclusion body formation that has implications for the methods used to



large family spanning three generations (results shown in the figure).

I administered fourteen tests, adapted from a battery of tests for aphasia¹, to almost all the members of the family, both normal and dysphasic (the former provided a control group), I also gathered data from samples of writing and from interviews. Not all language skills were equally impaired.

Four of the tests required aptitude with syntactical-semantic features. The responses of the dysphasic members of the family to those tests were significantly different from those of the normal ($P = 0.01$), but the responses of the two groups to the other tests were not significantly different. For several of the tests, the responses of 15 independent normals were available, but did not differ significantly from the responses of the normal family members.

By way of illustration, the dysphasics do not differ significantly from normals in their ability to point correctly to pictures instantiating reflexives ("He washes him" versus "He washes himself"), possessives ("The mother's baby" versus "The baby's mother") and negative passives ("The car is not being pulled by the truck"). But the two groups differ significantly in their ability to change tenses, to construct regular plurals for nonsense words, to detect a particular class of grammatical errors and to correct them.

For example, in a tense-changing test consisting of 10 items such as "Every day he kisses his nanny. Yesterday he —", the median score of the normals is 9, for the dysphasics, 3. Similarly, in a test of grammatical features consisting of 30 items such as "The boy eats three cookies", the median score for normals is

impaired. They may, for example, produce correct plurals for known words, but they lack a general rule for producing plurals. The dysphasics perform as normals in a task requiring that they point to "The book" rather than "The books", but when they are given a picture of an imaginary animal called a "wug", they cannot say whether a group of these animals would be called "wugs".

This and further evidence from the samples of writing suggests that the dysphasics, instead of being able to infer general rules about the signifiers of grammatical features from a few salient examples as normal children do, must learn each word as a separate lexical item. They have learned that the word "books" refers to several objects used in reading in much the way that normal speakers have had to learn that "children" refers to several young people. They appear not to know that there is a general feature of English grammar for signifying the number of nouns, determiners and verbs.

These findings are consistent with others. Clahsen², using data from the German language, and I (using data from refs 3 and 4) have shown that the language errors typical of dysphasia can be accounted for by the impairment of one particular grammatical faculty — the accurate usage of syntactical-semantic features of language such as the significance of number, gender, animacy, proper names, tense and aspect. Grammatically consequent skills are also impaired, as in the selection of apt determiners and the omission of subject pronouns before untensed verbs. Yet other grammatical skills, such as the judgement that the sentence "He puts" is ungrammatical, are unimpaired.

The deficits with which we are con-

purify and fold the proteins produced.

We have been producing Arg 52-BPTI, with methionine 52 replaced by arginine, as a fusion protein, releasing the BPTI moiety by cyanogen bromide cleavage at a preceding methionine residue, and purifying it in the reduced unfolded form. Our initial product appeared to be authentic and nearly pure, but precipitated in conditions conducive to disulphide bond formation and folding. At lower concentrations, the protein remained soluble but formed intermolecular mixed disulphide bonds rather than the correct intramolecular protein disulphide bonds. This behaviour was surprising; the folding behaviour of normal BPTI is well characterized² and we used conditions that normally produce quantitative refolding. The substitution of methionine 52 by arginine seemed an unlikely explanation.

A more plausible explanation is that the recombinant BPTI is isolated as a complex with some other tightly-bound substance that markedly affects its solubility and folding properties. BPTI is a very basic protein, whereas most intracellular macromolecules are acidic, so that the most likely candidate for a complexing partner would be a polyanion. Such an electrostatic complex might be most readily resolved by ion-exchange chromatography at extremes of pH, where the net charge on one component would be minimal. In accordance with this expectation, recombinant BPTI purified on a cation exchange resin at pH 2 was found to have the solubility and folding properties of normal BPTI. Our treatment seems to have removed an acidic component not yet identified. Another recent study³ has also demonstrated that the treatment of recombinant proteins with ion-exchange resins improves their solubility, but does not propose an explanation.

These observations suggest that inclusion body formation may sometimes occur because the newly synthesized polypeptide chain forms a tight complex with a component of the foreign cell in which it is synthesized, with unfavourable consequences for its solubility and folding properties. A variety of interactions may stabilize such complexes, so that there can be no general rules for dissociating them. Nevertheless, awareness of the possible presence of complexes should suggest ways of resolving them as well as the stage in the purification process at which refolding of the protein should be attempted.

NIGEL J. DARBY

THOMAS E. CREIGHTON

MRC Laboratory of Molecular Biology,
Hills Road,
Cambridge CB2 2QH, UK

1. Marston, F.A.O. *Biochem. J.* **240**, 1–12 (1986).

2. Creighton, T.E. *Prog. Biophys. molec. Biol.* **33**, 231–297 (1978).

3. Hoess, A., Arthur, A.K., Wanner, G. & Fanning, E. *Bio-Technology* **6**, 1214–1217 (1988).

Swallowtail performance

SIR—The claim¹ that genes for oviposition preference and genes for larval performance are not linked in the host specific swallowtail butterflies *Papilio zelicaon* and *P. oregonius* would have important implications for the evolution of phytophagous insects. But the conclusion seems to be premature.

Oviposition preference in these species is controlled by genes on the X chromosome², which is why Thompson *et al.*¹ sought evidence that larval performance is also sex-linked. Since females are heterogametic in butterflies, sex linkage predicts that female F1 hybrid larvae should perform like their male parents. There is clear evidence from survivorship of larvae that *P. zelicaon* and *P. oregonius* perform better on their own host plants, but survival could not be used to test sex linkage because male and female larvae could not be distinguished. Two other measures of larval performance were used instead — development time and pupal mass of surviving larvae — that could be measured on

the two sexes separately. No evidence for sex linkage was found.

The problem with these two measures of performance is that they do not appear to be good indicators of adaptation to the host plants. For both measures, *P. zelicaon* has a lower score than *P. oregonius* on both hosts. The difference is somewhat greater on the natural host of *P. oregonius*, but this effect is slight and no significant host/species interaction is quoted by the authors. If the characters scored are not good measures of adaptation to alternative hosts, then the absence of sex linkage is irrelevant to the question of genetic linkage between oviposition preference and larval performance.

ROGER BUTLIN

Department of Pure and Applied Biology,
University of Wales College of Cardiff,
PO Box 915,
Cardiff CF1 3TL, UK

1. Thompson, J.N., Wehling, W. & Podolsky, R. *Nature* **344**, 148–150 (1990).

2. Thompson, J.N. *Evolution* **42**, 1223–1234 (1988).

Glucose in glucagon release

SIR—We read with interest the contribution to Scientific Correspondence by Gylfe¹ concerning our recent Letter². The concept that glucose lowers $[Ca^{2+}]_i$ has recently been proposed by Gylfe and co-workers to be the mechanism by which the sugar inhibits glucagon release^{3,4}. Although we cannot disregard such a mechanism, the existing experimental support is insufficient. First, all experiments were performed on cells at a basal glucose concentration of 0 mM, suggesting that the α_2 cells were deprived of fuel. Consequently, when the glucose concentration was restored to physiological resting levels (about 5 mM) the ATP-dependent Ca^{2+} pumps can be expected to be activated with a resultant lowering in $[Ca^{2+}]_i$. Second, Gylfe and co-workers have not shown that glucagon is present in the cells subjected to $[Ca^{2+}]_i$ measurements. Finally, in their studies, changes in $[Ca^{2+}]_i$ were not correlated to glucagon release.

For the neurotransmitter GABA co-secreted with insulin to be an effective regulator of glucagon release, it should have a pronounced effect already in the early phase of insulin release. Gerich *et al.*⁵ found an increase in insulin and thereby GABA release at about 2.5 mM glucose, consistent with the idea that GABA inhibits glucagon release. Further, as we pointed out², the GABA_A-receptor antagonist bicuculline counteracts the inhibitory effect of glucose on glucagon release. These data, together with the observation of glucagon hypersecretion in non-insulin-dependent diabetes mellitus, support our hypothesis².

Although our hypothesis does not exclude other mechanisms by which glucose may inhibit or modulate glucagon release, it is probably the chief mode of action of the sugar. Indeed, it is of interest that a second level of glucose control of glucagon release may be possible by direct metabolic modulation of the GABA_A receptor, as has been recently demonstrated⁶ in dissociated mouse neurons.

With respect to citation, we did not state that Okada *et al.*⁷ were the first to demonstrate GABA within the pancreatic islets. That GABA is present in the endocrine part of the pancreas at concentrations comparable to those encountered in the central nervous system, however, is highly relevant to the credibility of our hypothesis.

P. RORSMAN

Department of Medical Physics,
Gothenburg University,
Box 33031, S-400 33 Gothenburg,
Sweden

P-O. BERGGREN

Department of Endocrinology,
Karolinska Institute,
Box 60500, S-104 01 Stockholm,
Sweden

P. A. SMITH

University Laboratory of Physiology,
Parks Road, Oxford OX1 3PT, UK

1. Gylfe, E. *Nature* **344**, 300 (1990).

2. Rorsman, P. *et al.* *Nature* **341**, 233–236 (1989).

3. Johansson, H., Gylfe, E. & Hellman, B. *Biochem. biophys. Res. Commun.* **147**, 309–314 (1987).

4. Johansson, H., Gylfe, E. & Hellman, B. *Cell Calcium* **10**, 205–211 (1989).

5. Gerich, J.E. *et al.* *A. Rev. Physiol.* **38**, 353 (1976).

6. Duchon, M.R. *J. Physiol.* **415**, 48 (1989).

7. Okada, Y., Tamaguti, H. & Shimada, C. *Science* **194**, 620–622 (1979).

Founders' feast

Ralph A. Lewin

Handbook of Protoctista. Edited by Lynn Margulis, John O. Corliss, Michael Melkonian and David J. Chapman. *Jones and Bartlett: 1990. Pp. 1,024. \$195. Distributed in the United Kingdom by Van Nostrand Reinhold, £195.*

If I had written this review in Esperanto — a perfectly phonetic international language, much more logical in construction than English — I doubt whether very many readers of *Nature* would understand it. Verbal communication entails a common understanding of words used by the writer for the readers. If a word means one thing to the writer and something different to the reader, then understanding is impaired. Now consider the terminology of this great book: does it help or hinder understanding? The title, for instance: what are the Protoctista? They are, literally, the founding fathers: proto=first + ctistes=founders. (Note that the accent goes on the *ctist*, not on the *toc*.) The word, according to the senior editor, was coined by a Scot, J. Hogg, in 1861, and disinterred by Corliss and Margulis in recent years. Margulis suggests that protoctistologists should now secede from

the zoological and botanical unions and institute their own international protoctistological type collection. (I wish them luck!)

The Protoctista comprise what most biologists know as unicellular animals or protozoa, along with unicellular and multicellular algae and some fungi. They have been reassigned here to about 35 phyla, mostly new. Here, seaweeds are not plants, because by the editors' definitions plants are identified with bryophytes + tracheophytes. Although some seaweeds are algae — the brown ones, for instance — the red algae are classified with *Dictyo-stelium* and the Conjugales because they lack undulipodia.

These organelles are more usually called flagella or cilia, but Margulis prefers the bastard word (half-Latin, half-Greek) because eukaryote flagella are so very different in size and structure from

those of bacteria. (By the same reasoning, we should use different words for the wings of birds and insects which, though both are organs of flight, differ fundamentally in size, embryological origin and construction.) The word undulipodia (I assume in Cyrillic) was used in a Russian text, Margulis tells us, by A. P. Shmagina in 1948; it has been transliterated and used by at least two of this book's editors.

Personally, I prefer the term "enneane-ma", which is more euphonious, uniformly Greek in origin, and reminds one of the nine-stranded construction characteristic of these organelles, rather than suggesting that they are undulant (which some are not) feet (which they certainly aren't). For Margulis, undulipodia are intracellular (because they are inside a membrane), whereas the flagella of bacteria and the axial fibrils of spirochaetes are extracellular (because they are outside the inner plasma membrane).

Not everyone would agree with such distinctions. Our old schoolfriend *Spirogyra*, in this treatise, is not a Chlorophyte, nor is *Saprolegnia*, which tends to infest fishes in our home aquaria, a fungus. (The kingdom of fungi is outside the realm of this treatise.) It's all likely to be very confusing. Margulis, like Humpty Dumpty in "Alice in Wonderland", wants words to mean what she decrees for them, whereas I prefer to use words that are, as far as possible, immutably understandable.

There is a more serious element to my critical comments. We are experiencing, in these days of blooming molecular biology, radical revisions of our knowledge of phylogeny, particularly as we become aware of anastomoses among the branches. (Margulis, stating that "all protoctists descend from tightly knit bacterial communities", refers to growing evidence that plastids and mitochondria had symbiotic prokaryotic ancestors.) Nucleotide sequences, and hence amino-acid sequences, are enabling us to construct phylogenetic trees that are, we trust, increasingly objective indications of the pathways of evolution, and our taxonomies will ultimately have to recognize the results of these new disciplines. Just as it has been said that translations of poetry are like mistresses — they can be either beautiful or true, but rarely both — so classifications can be based either on molecular-biology phylogenies, or on human convenience, though these may not be reconcilable. Personally, as we have to put specimens into cabinets and relevant information into chapters, I lean conservatively towards convenience at the present time. *Paramecium* may be, in margulese, an undulipodiated protoctist, but for me it remains simply a ciliate.

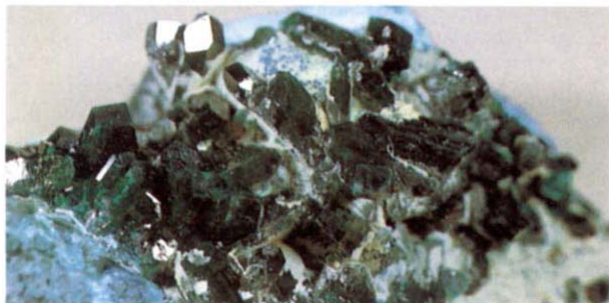
Margulis' real forte is evident in the conceiving and organizing of this, the first encyclopaedic volume to embrace algae, aquatic fungi and protozoa. It has been



Linarite



Crocoite



Dioptase

Linarite, Crocoite and Dioptase are just three of the minerals studied in *Rocks, Minerals and Fossils of the World* by C. Pellant. Minerals are shown mostly as typical specimen material rather than perfect museum exhibits, thereby aiding potential matching of samples. Each mineral that is represented is accompanied by a full description of its properties and some indication of its distribution, together with notes on its more unusual modes of occurrence. Rocks are divided into three sections: igneous, metamorphic and sedimentary, with 'field' photographs illustrating the relationship between strata and rock content. Fossils are arranged in biological groups and stratigraphically. The volume is beautifully illustrated by the photography of R. Phillips and C. Pellant. Published by Pan Books, price £12.99. Distributed in the United States by Little Brown, \$29.95.

almost a decade in the making. With her three co-editors, Corliss, Melkonian and Chapman, she has marshalled a formidable team of about 60 collaborators, who have reviewed their special fields or phyla in a more or less uniform format. They were supported by funds from four organizations, aided by more than 30 helpers, and used at least four computer programs: all these are generously acknowledged.

Some phyletic distinctions seem on shaky ground. For instance, it is asserted that complex sexual cycles occur among the Xanthophyta but not among the Eustigmatophyta, although one may question whether such distinctions will hold up with further study. And I'm far from sanguine about the survival into the next millenium of many of the lesser phyla like the pyronymphids (metamonads), bicoecids, haplosporidians and xenophyophorans.

Each chapter, limited to 40 pages or less, consists of a survey of the organisms in that phylum, their ecologies and taxonomies, with a few details, when known, on cultivation techniques and evolutionary history. Many are lavishly illustrated with electron micrographs, as well as photographs, drawings (too many, alas, lacking scales), diagrams and charts. The 80-page glossary comprises around 4,000 definitions, mostly useful (clade, cline, clone), but quite a few of questionable relevance (chalk, chamber, chert).

I feel that the creation of such terms as "detrivory" and "bactivory" betrays a lack of feeling for classical roots. And are such long names as Postciliodesmatophora and Plasmodiophoromycota really necessary? They are surely not convenient to use. Do microscope slides or museum jars bear them? Do experts, chatting over the phone, use them? Or are they just exercises in word construction for someone's attempted aggrandisement? There are also three-column indexes of organisms (48 pages), most of the cited authors (25 pages) and general subjects (47 pages): who could ask for more?

The model is Bergey's *Manual of Systematic Bacteriology*; like it, this is indubitably a milestone (and, at 2.5 kilograms, a weighty one). I have not critically read all of it; I couldn't, and I doubt if anyone could. But to judge by its treatments of the algae in the classical sense, which occupy about a third of the work, this monograph is certainly authoritative and doubtless will find its way on, and frequently off, the shelves of biologists everywhere. Although it is apparent that few of the contributors were happy to accept all the editors' terminological neologisms, the book is no whit the worse for such slight inconsistencies. Its title notwithstanding, I can see that I shall be using it a lot. □

Ralph A. Lewin is at the Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA.

The long-range forecast

Alan J. Thorpe

Storm and Cloud Dynamics. By William R. Cotton and Richard A. Anthes. *Academic*: 1989. Pp.883. \$149.95, £107.50.

THE International Geophysics series is a set of monographs and textbooks in the field of solid Earth geophysics, atmos-

precipitation, fall within the scope of this book. This notion of a distinction in terms of scale is a difficult one and can distract even the specialist.

This book is a research monograph and not a teaching text. The authors cover the literature of their subject in encyclopaedic detail, with only rare lacunae. The emphasis in the title is on dynamics, but this is not adhered to strictly. Within meteorology the qualification 'dynamics' means a description of processes involving air motion rather than those solely to do with physical processes such as water-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Cumulus storm clouds and rain, Kalahari, Gemsbok national park in Botswana.

pheric physics, oceanography and meteorology. The volumes on meteorology consist of several important teaching texts — such as Gill's *Atmosphere and Ocean Dynamics* and Holton's *Introduction to Dynamic Meteorology* — as well as classic monographs such as Palmén and Newton's *Atmospheric Circulation Systems: Their Structure and Physical Interpretation*.

Storm and Cloud Dynamics is a title needing some clarification, particularly after the recent unusually severe weather in the United Kingdom. The term 'storm' is used in popular and in meteorological parlance to denote any weather event that can be regarded as violent, particularly wind and rain or snow. But the authors here use the word 'storm' in a narrower sense to refer to precipitating mesoscale weather systems. Mesoscale delineates those horizontal scales of atmospheric flow between several kilometres and several hundred kilometres, whereas precipitating refers to rain, hail or snowfall. So, for example, the 'storms' affecting the United Kingdom in February do not directly qualify as they have a scale of a few thousand kilometres. Of course, even large cyclonic storms such as these have mesoscale sub-structures embedded within them which, if they relate to clouds and

vapour condensation or radiative transfer. This distinction is, anyway, artificial and the authors of this text, in not confining themselves to a narrow definition of dynamics, are modern in their outlook. The penalty for this modernity is length. Further, the field of mesoscale meteorology is rapidly developing; new observations and theories abound, making the writing of a monograph a challenging task. Weather-prediction computer models now have a resolution sufficient to begin to resolve these mesoscale structures, and exciting new observational technologies are revealing hitherto unknown phenomena. So the field is at the forefront of meteorological science, and this monograph is thus invaluable to researchers.

Mesoscale weather systems include fronts, intense mid-latitude cyclones, tropical cyclones, cumulonimbus clouds, tornadoes, mountain windstorms, polar lows and mesoscale convective complexes. Not all of these are described here as the focus is on those systems with clouds. Significant advances have been made in simulating individual thunderstorm clouds using computer models with gridpoints spaced at perhaps one-kilometre intervals. These are not weather-forecast models but rather models representing

basic processes. Such modelling results are described comprehensively here. Researchers at the new Center for the Analysis and Prediction of Storms in Oklahoma plan to construct a forecast model capable of predicting the development and evolution of thunderstorms day-by-day. Until that enterprise is successful the most advanced operational mesoscale forecast model, at the UK Meteorological Office, has a horizontal resolution of about 15 kilometres.

Within the field defined by the authors few stones are left unturned. The style of the book is one in which the body of knowledge on a given topic is summarized comprehensively, which makes the volume a storehouse of invaluable information. It does not, on the whole, present the authors' viewpoint or perspective. This is not to say that the authors have not contributed significantly to research in mesoscale meteorology. They have, and they make reference to their work, but

their vision of the subject is not clearly evident here.

I am sure that this book will be read by all mesoscale meteorologists, and particularly by graduate students and people about to enter the field. Of course, one can quibble about what is included and what glossed over. The sections on cumulonimbus dynamics and on mesoscale convective systems are good, whereas that on frontal rainband dynamics is cursory and misleading. One might question the need for so much on airflow over mountains, cloud parameterization and radiative transfer in a book on cloud dynamics.

In conclusion, the authors provide a scholarly account of a large, actively developing topic. The result is immensely useful. □

Alan J. Thorpe is in the Department of Meteorology, University of Reading, Whiteknights, Reading RG6 2AU, UK.

Cashing chips

Richard L. Gregory

The Industrialization of Intelligence: Mind and Machine in the Modern Age. By Noah Kennedy. *Unwin Hyman*: 1989. Pp. 213. £14.95, \$24.95.

STARTING with the great library in Alexandria and its tragic loss, and ending with an analysis of the role of computers in present-day society with speculations for the future, this book is an essay on the significance of mechanized intelligence from papyrus scrolls to silicon chips. In the middle we meet artificial intelligence: whether brains are intelligent computers, and whether computers can be intelligent brains. This is a fascinating question which may be thought of as purely empirical — is it actually possible to make fully intelligent computers? — or as philosophical. John Searle argues on philosophical grounds that computers can handle only syntax, without appreciating meaning. He argues this by analogy with a closed room in which Chinese symbols are moved around correctly though without understanding. Kennedy's conclusion is that what Searle's Chinese room argument "has really done is to muddy the crystalline waters of the strong AI approach". One might indeed agree, but even if it has not provided direct light the debate at least has inspired some productive heat. The interest is to follow the debate; but this requires a depth of description and analysis of sometimes technical issues. These are generally lacking here, which makes *The Industrialization of Intelligence*, despite its dramatic and important theme, curiously featureless

and flat.

In this book one does not find definitions of intelligence, or of creativity, and comparisons between human and machine 'mental' performance are sadly lacking. Some pictures would have helped. Why have a long detailed description of the layout of a Babbage computer when a line drawing or photograph would have conveyed it clearly at a glance?

The assemblage of scholarship on some significant issues may be useful, especially perhaps for planning future industrialization of intelligence. One is reminded that at the start of the computer revolution the British government estimated that but two computers, at most, would be required. □

Richard L. Gregory is in the Department of Psychology, University of Bristol, Bristol BS8 1HH, UK.

Spring Books

NEXT week is *Nature's* Spring Books issue. Books to be reviewed include the autobiography of Hans Eysenck and the second volume of the collected papers of Albert Einstein. Daniel S. Greenberg discusses the relationship between scientists and the government agencies that support them. David Owen, leader of Britain's Social Democratic Party, analyses a prediction of the problems that will face the United States and Japan, in particular, by the end of the century. There are stories of explorers: Admiral Peary's disputed claim to have reached the Pole is reassessed in the light of new 'scientific' evidence, and the story of the exploration of the Amazon is told. There are accounts of the history of science: medical research in Britain and the development of the field of biological oceanography. The monumental work *The Ants*, by B. Hölldobler and E. O. Wilson, is reviewed, and on a lighter note, Walter Gratzer turns his attention to that indispensable accessory, the pencil. □

Out of sight

Stuart Sutherland

Neuropsychology of Visual Perception. Edited by Jason W. Brown. *Erlbaum Associates*: 1989. Pp. 267. £29.95, \$49.95.

THE following story about the late Professor Oliver Zangwill illustrates the complexity of the effects of brain damage. Zangwill once investigated a patient with severe visual agnosia, that is the inability to recognize objects by eye. At the beginning of the first session, Zangwill showed his own magnificent blue fountain pen to the patient and asked him to try to remember it. When Zangwill presented it again at the end of the session, the patient completely failed to recognize it. This procedure, with the same pen and with the same result, was repeated for ten sessions. At the end of the final session, when the patient yet again failed to recognize the pen, Zangwill, at his wits' end, asked, "Do you know who I am?". "Of course", said the patient unhesitatingly, "you're the man with all those fountain pens".

Until fairly recently, work on brain-damaged patients was directed almost wholly at discovering in which part of the brain a particular function is carried out. Nowadays, neuropsychologists believe that investigating the differential loss of different functions may shed light on how the normal brain works. For simple functions, this approach has sometimes produced results that are readily interpretable. For example, patients may lose colour vision while retaining the ability to see motion, or vice versa. In fact little was made of this finding until it was discovered by microelectrode recording that colour vision and motion perception are subserved by different areas of the visual cortex. Unfortunately, neuropsychologists have not helped to uncover the brain mechanisms that underlie either function: for that we have to turn to psychological work on normal people and to recent neurophysiological findings with microelectrodes.

To take more complex abilities, *Neuropsychology of Visual Perception* contains evidence from brain-damaged patients that object recognition and route finding are also dissociable. But when one reads the comprehensive and informative chapter by M. J. Riddoch and G. W. Humphreys on the latter topic, it becomes obvious that this ability can itself break down in many different ways, even though most of the lesions that give rise to it are in approximately the same part of the parietal lobe. Some patients can continue to find their way around a familiar environment, but cannot read a map or draw a plan; for others these defects are reversed. In some cases, depth vision is

deficient and in addition to their topographical impairment, patients can bump into objects quite violently. Others can see parts of objects but cannot see the relationship between the parts. The authors suggest that another patient was impaired because she could not manipulate enough information at a time to plan a route or to keep track of it once embarked upon. Because it is obvious that all these abilities must enter into route finding, the neuropsychological findings do not seem to be telling us anything more than that each ability can be selectively impaired by lesions.

I have touched on only one of the many defects dealt with in this book. Others include unilateral neglect, in which a lesion to one hemisphere can cause the patient to be unaware of half of his visual field: he may eat the food that is on the same side of the plate as the damaged hemisphere while leaving that on the

other side untouched; when asked to describe a scene from memory, he may describe only that part of it lying on the same side as the lesion. Another strange phenomenon is palinopia: the image of something seen may persist so that when a hat is seen, the same hat subsequently appears on different people's heads.

J. W. Brown has edited a useful collection of papers that will for some time form a standard work of reference for anyone interested in the field. But, despite the fascination of the bizarre defects described, they have so far failed to illuminate the workings of the visual system. They remain more of a challenge to explanation than a contribution to it, for like Zangwill's patient, they raise more problems than they solve. □

Stuart Sutherland is in the Department of Experimental Psychology, University of Sussex, Falmer, Brighton BN1 9QG, UK.

Model behaviour

Grant R. Bigg

El Niño, La Niña, and the Southern Oscillation. By S. George Philander. *Academic: 1990. Pp.289. £42.50, \$59.50.*

THE trade winds in the central equatorial Pacific weakened considerably last year. Convective activity along the Equator near the dateline strengthened as 1989 passed into 1990, while sea surface temperatures were 0.5 to 1 °C above average along the Equator between 160 °E and 150 °W during January. Nevertheless, forecasts supplied to the Climate Analysis Center in Washington, DC, suggest that climatic conditions during 1990 will not be unusual, and specifically that El Niño is not about to disrupt the Pacific basin once again.

How have we arrived at the situation where dynamical models of the atmosphere and oceans can be used with moderate confidence to infer interannual climatic changes over a significant portion of the globe? George Philander in his new book captures the story behind the present embryonic science of forecasting climate variability, and comprehensively explores the coupled physics of the atmosphere and oceans that forms the basis of this science. Much knowledge has been gained recently. A major El Niño event occurred in 1982, causing great economic and climatic hardship in the Pacific basin and elsewhere. There were droughts in Australia, India and southern Africa; floods or typhoons soaked Ecuador, California and Tahiti; and the anchovy-fishing industry off Peru was practically destroyed. This quasi-periodic, large-

scale alteration to the distribution of heat and momentum in the ocean and atmosphere of the tropics, particularly of the Pacific, took the scientific community by surprise because of its timing and intensity. Such clear evidence of a lack of understanding of a phenomenon that semi-periodically affects a quarter of the globe led to a focusing of research efforts onto the physics of the coupling of the atmosphere and ocean.

Philander leads the reader on a voyage of discovery through the known climatology and variability of the atmosphere and oceans of the tropics. He demonstrates clearly the complexity of the system that needs to be understood, exploring the mathematical and physical models that have been constructed to explain the ocean and atmosphere systems in isola-

tion. Finally, he presents the recent advances in the understanding of the coupling of these two systems, through the use of simple models.

The dominating message of these descriptions, whether of the ocean, the atmosphere or the coupled system, is the ubiquity of waves. Whether one is considering the response of the ocean to an applied wind stress, the atmospheric heating from a sea surface temperature anomaly or the interaction of the two media over entire ocean basins, wave processes are vital to physical understanding. Particularly in trying to unravel the coupling of the ocean and atmosphere, we are shown clearly how the theoretical exploration of different modes of climatic oscillation, begun only in the past few years, has opened up possibilities of forecasting interannual climatic variability. These possibilities are currently being investigated through the development of coupled general circulation models of the ocean and atmosphere.

This book is not for the casual reader and should be approached only by those with solid physical and mathematical skills. But it illuminates a field that has intrigued many since Sir Gilbert Walker, director-general of observatories in India in the early part of the century, demonstrated strong trans-Pacific correlations in climate. Philander has had considerable personal experience in the search for explanations of tropical climate variability. The synthesis of this experience shows that much still needs to be understood but also demonstrates the knowledge gained since 1982 and the hope that future El Niños will not catch the world unprepared. □

Grant R. Bigg is in the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.



Long in the tooth — the walrus is one of the marine mammals described in *The Pinnipeds* by M. Riedman. Published by University of California Press, price \$29.95, £18.70.

Sex determination compared in *Drosophila* and *Caenorhabditis*

Jonathan Hodgkin

Fruitflies and nematodes show many similarities in the general organization of the gene networks that control sexual dimorphism and dosage compensation. In contrast, the underlying molecular mechanisms appear to be very different in these two species. Developmental processes such as sex determination need not be strongly conserved in evolution.

THE study of animal development by genetic methods has tended to concentrate on a very limited number of species, of which the two most conspicuous are the fruitfly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*. These two animals are phylogenetically remote from each other, and differ in many aspects of their biology and development. Comparison is consequently difficult. However, only through comparative studies is it possible to identify developmental strategies that are conserved during evolution, as opposed to those that are confined to particular taxonomic groups.

One area in which extensive comparisons can be made is the generation of sexual dimorphism. Sex determination and sexual differentiation are processes that are almost universal in the animal kingdom, and they have been studied in great detail in both nematode and fruitfly. At a superficial level, sex determination is remarkably similar in these two species. However, we now know enough about the underlying genetic and molecular mechanisms in the fly and the worm to conclude that the general similarity in overall organization is *not* reflected at more mechanistic levels. The similarity in organization must therefore be the result of convergence. The broad similarities between the two systems are set out in Fig. 1 and Table 1a; the differences are set out in Table 1b.

The present state of knowledge of sex determination in the fruitfly and the nematode is the result of combined genetic and molecular approaches. Many mutants altered in sexual phenotype have been generated in both species (some examples are given in Table 2), permitting the identification of key regulatory genes. Extensive genetic analysis has led to detailed models for the roles of different genes and the interactions between them. However, the genetic models make few predictions about the molecular nature of the interactions between the genes, or whether these interactions are direct or indirect. Many of the genes discussed in this article have now been cloned, so answers to these questions are becoming available. The molecular data have so far supported the genetic predictions, but have also revealed unexpected phenomena.

The purpose of this review is twofold: first, to outline the two systems and indicate the ways in which the study of sex determination touches on many problems of general biological importance; second, to discuss the similarities and differences between the two species. Each of the areas discussed has been the subject of a substantial body of work, so this account is necessarily superficial.

Several recent reviews provide more detailed coverage of specific areas (for *D. melanogaster*, refs 1–5; for *C. elegans*, refs 6–10).

Overview

In both these organisms, the primary sex-determining signal is the ratio of X chromosomes to autosomes. In one sex, there are two sets of autosomes and two X chromosomes (2X; 2A, ratio 1.0) and in the other there are two sets of autosomes and one X chromosome (1X; 2A, ratio 0.5). This primary signal acts to set the functional state of a collection of major regulatory genes, which govern three largely independent processes: dosage compensation, somatic sexual development, and germ-line sexual development (Fig. 1).

Dosage compensation is the process whereby total gene expression from the single X in one sex is made equal to total expression from the two X chromosomes in the other sex. This can be achieved either by up-regulating the single X in one sex, or down-regulating the two X chromosomes in the other, or by a combination of these mechanisms. Failure to establish correct dosage compensation is usually lethal in both *D. melanogaster* and *C. elegans*. Somatic sexual development refers to sex-specific differentiation in all tissues except the germ line. Germ-line sexual development refers to the type of gamete made by the germ line, either sperm or oocytes. Major dimorphic features in the two species are indicated in Figure 2.

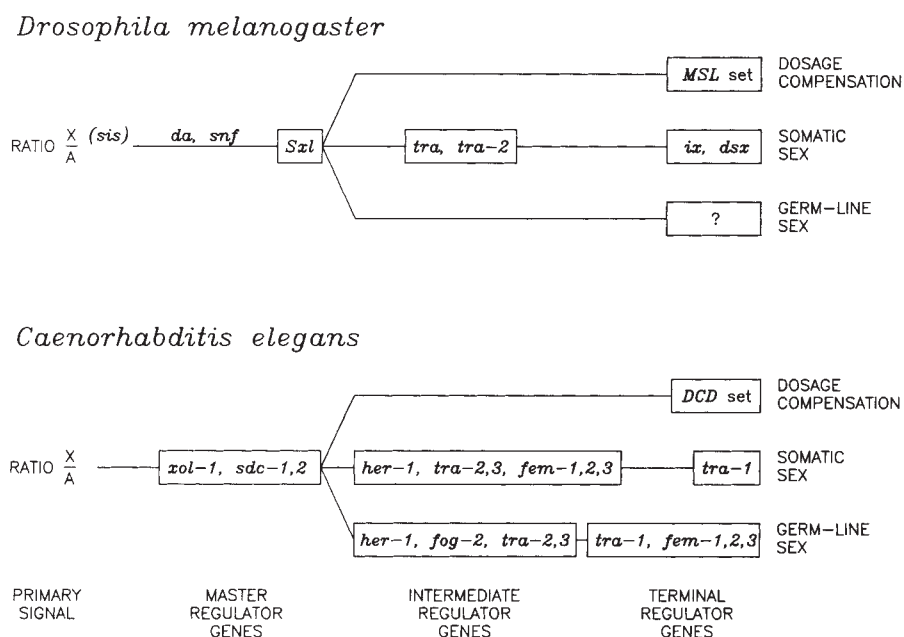
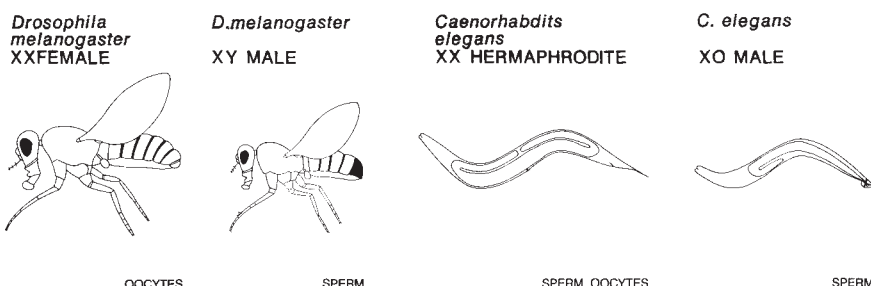


FIG. 1 General organization of sex determination gene networks in *D. melanogaster* and *C. elegans*. Genes with major regulatory roles at specific points in the two systems are indicated. The regulatory events within each network are set out in greater detail in Figs 3–8. Various genes with uncertain roles have been omitted, for example *her* and *ovo* in *D. melanogaster* (see refs 1, 57), and *dpy-29* and *fog-1* in *C. elegans* (see refs 8, 60).

FIG. 2 Sexual dimorphism in *D. melanogaster* and *C. elegans*. These very schematic diagrams indicate some of the major sexually dimorphic features of the two species. In *D. melanogaster*, somatic sex differences include major differences in the genitalia and internal gonadal structures. External features that are easily scored are the abdominal pigmentation and the male-specific sex-comb on the fore-leg. In addition, females are larger and there are many internal differences, such as the female-specific synthesis of yolk proteins in the fat-body and ovary. The nervous system must contain many sex-specific specializations in order to generate the distinctive male and female courtship patterns. In the nematode, there are extensive differences in the soma⁶. About one quarter of the somatic nuclei in the male are involved in constructing the specialized copulatory organ in the tail. Features that are easily scored are the organization of the gonad (two-armed in hermaphrodites, one-armed in males), the presence of a centrally-located vulva (in hermaphrodites) or copulatory tail structures (in males). Additional differences include the larger size of hermaphrodites, hermaphrodite-specific synthesis of yolk



In *D. melanogaster*, the natural sexes are the XX female and the XY male. The *Drosophila* Y chromosome is not male-determining (as it is in mammals), and there are close relatives of *D. melanogaster* such as *D. annulimana* that have no Y chromosome and consequently show an XX female/XO male arrangement¹¹. An XX female/XO male system is also found in most nematodes, including close relatives of *C. elegans*^{11,12}.

C. elegans itself shows a modification of this basic arrangement, because the natural sexes are the XX self-fertilizing hermaphrodite and the XO male. It is likely that the hermaphrodite sex of *C. elegans* represents a secondary specialization, and that the hermaphrodite should be regarded as a female that is able to make a limited number of sperm before switching over to sustained oogenesis. Genetic analysis has strongly supported this view of the hermaphrodite as purely female in its soma, and mixed (first male, then female) in its germ line. Part of the interest in studying sex in *C. elegans* is to understand the way in which a mixed-sex germ line can be achieved.

In both organisms, therefore, there is a dichotomous choice between male and female development in each of the areas

proteins (in the intestine) and many well-defined differences in neuron number and connectivity (hermaphrodite nervous system 302 neurons, male nervous system 381 neurons). In *C. elegans* adults of either sex, at least half the nuclei belong to the germ line. These are exclusively destined for spermatogenesis in males, but in hermaphrodites the early gametes differentiate into sperm (approximately 300 are made), and all subsequent gametes differentiate into oocytes.

under discussion. This choice is ultimately determined by the X/A ratio in the developing zygote.

The X/A ratio

Classical experiments long ago¹³ established that sex in *D. melanogaster* is established by the ratio of the number of X chromosomes to the number of sets of autosomes, or X/A ratio for short. Varying the total ploidy level made it possible to examine the effect of ratios other than the natural 0.5 and 1.0. These experiments made it clear that it was the ratio, rather than the absolute number of X chromosomes, that determined sex. The effects of altered ratios are summarized in Fig. 3.

Corresponding experiments in *C. elegans* (also summarized in Fig. 3) revealed almost the same pattern: X/A ratios less than 0.67 resulted in male development, ratios greater than 0.75 resulted in hermaphrodite development¹⁴. In each organism, therefore, a sharp threshold is in operation. This is one of the best examples of a threshold response in developmental biology, and raises the question of how the embryo is able to compute this ratio so accurately and reliably.

In both organisms there seem to be multiple factors on the

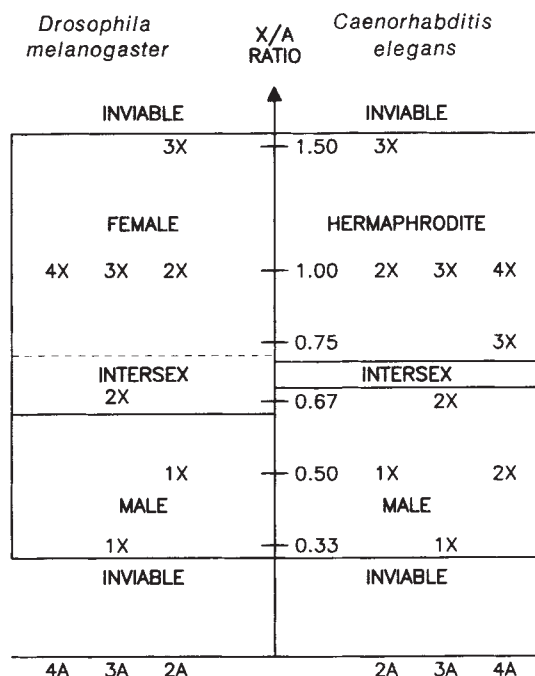


FIG. 3 Effect of X/A ratio on sexual phenotype. Different values of the X/A ratio are plotted vertically, with their effects on phenotype in *D. melanogaster* to the left, and *C. elegans* to the right. These different values are generated experimentally by varying the number of X chromosomes from one to four in the presence of diploid (2A), triploid (3A) or tetraploid (4A) sets of autosomes. At ratios too low (<0.3) or too high (>1.5) the dosage compensation mechanism is unable to prevent death. Within the viable range, low ratios lead to male development and high ratios to female (fly) or hermaphrodite (nematode) development. The threshold range, over which intersexual development ensues, is slightly different in the two species, because a ratio of 0.67 leads to intersexual development in flies but male development in nematodes.

X chromosome that contribute to this ratio. Equivalent experiments^{14,15} on fruitflies and nematodes showed that partial duplications of the X chromosome that had no effect on a diploid single X male (1X; 2A, ratio 0.5) could nevertheless significantly feminize a triploid 2X; 3A animal (ratio 0.67). By this assay, many different regions on the X chromosome in each species seemed to have a feminizing effect. This implied the existence of *multiple numerator elements*, which contributed additively to the X part of the X/A ratio. Whether these elements corresponded to chromosomal sites or to genes specifying gene products was not clear.

These observations on triploid animals may have been misleading, in that they implied the existence of a large number of numerator elements. There may be only a few, at least in *D. melanogaster*. The experiments are compromised by the involvement of dosage compensation effects, which can interfere in at least two ways. First, 2X; 3A flies are intersexual in phenotype, consisting of a mosaic mixture of male and female cells. Increasing the amount of X chromosome material by means of a partial X duplication will have a selectively deleterious effect on the male cells, because they have adopted the male (high) mode of dosage compensation. This is intended for operation at a ratio of 0.5, and is therefore increasingly deleterious at higher ratios. Consequently the mixture of male and female cells is shifted in favour of the female cells, and apparent feminization is seen although no numerator elements may in fact have been involved¹.

In recent years, sophisticated genetic experiments have sought to define numerator elements more directly, by their interaction with the target of the fruitfly X/A ratio, the gene *Sxl* (*Sex-lethal*). These searches suggest that there may be few major numerator elements¹⁶. Two elements, *sis-a* and *sis-b* (*sisterless*), have been defined so far. The two elements seem to be related but not identical, because a defect in *sis-a* can be partly but not wholly corrected by a duplication of *sis-b*.

The second potential problem with experiments on X/A ratio applies to the nematode. Dosage compensation in *C. elegans* seems to work by a negative mechanism (discussed below), so that a specific set of genes acts in XX animals to reduce the level of X chromosome transcripts. Consistent with this is the observation that in XX animals, partial X chromosome duplications lead to elevated expression from regions of the X that have not themselves been duplicated¹⁷. The interpretation is that the additional X chromosome material is titrating out putative general negative regulators of X chromosome expression. It could therefore be that only a few genes on the nematode X chromosome are contributing to the numerator signal, but that their expression level can be increased by duplications of other parts of the X chromosome. If the assessment of X/A ratio involves measurement of gene product levels rather than of chromosomal sites, then the effective X/A ratio will also be increased by these duplications.

Recent work on both organisms has opened up avenues to the molecular definition of numerator elements. In *D. melanogaster*, an important discovery¹⁸ is that the element *sis-b*

corresponds, at least in part, to the T4 transcript of the achaete-scute complex (AS-C). The AS-C genes encode proteins that are probably transcription factors, since they are members of the *myc* family^{19,20}. Their main role is in neurogenesis, and their main phase of transcription is relatively late in embryogenesis. However, there is also an early burst of general AS-C transcription, early enough to act in the establishment phase of sex determination²¹. It may be that the numerator elements are simply a small number of sex-linked transcription factors, synthesized before dosage compensation is established. The resulting higher level of these factors in XX embryos could be all that is needed to activate *Sxl* embryonic transcription (discussed below).

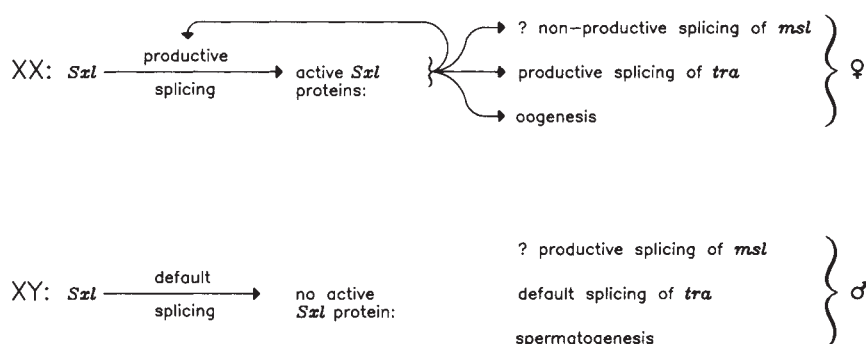
In the nematode, direct microinjection²² of cloned X chromosome DNA into polyploid animals has led to the identification of sequence elements that are able to shift the sexual phenotype of 2X; 3A; *mnDp8* animals from male to intersexual or hermaphrodite (*mnDp8* is a small X duplication, so the ratio is about 0.7). Some, but not all of the X chromosome clones tested were effective in this assay, and autosomal DNA clones tested were all ineffective. Deletion analysis of two different clones defined small non-coding regions that were sufficient to cause this shift in phenotype, and sequence comparison revealed a sequence of eight bases held in common between the two. This sequence seems to be present many times on the X chromosome, consistent with the expectation for a multiple numerator element. Experiments of this type have thus far given only transient and variable results, perhaps because the injected material is present as an unstable extrachromosomal array. Moreover, the defined elements appear ineffective at a lower X/A ratio (0.67). It remains possible that these sequences correspond to dosage compensation elements (binding sites for suppressors of X chromosome expression), rather than to true numerators, for the reasons explained above. Alternatively, there may be several qualitatively different kinds of numerator (as in fruitflies), and these experiments are, as yet, only looking at one kind. If these abundant short non-coding sequences on the nematode X chromosome do correspond to numerator elements, then the X/A ratio must be computed in rather different ways in nematodes and fruitflies.

In summary, both organisms appear to have more than one numerator element (but possibly not a large number), and some of these elements may soon be well defined in molecular terms. 'Denominator elements' constituting the autosomal part of the X/A ratio have not yet been defined in either flies or worms. Quite possibly they do not exist in the same specialized sense as *sis-a* and *sis-b*, but instead correspond to the large number of autosomal genes encoding the general transcriptional machinery of the cell.

Sex-linked master control genes

In both nematode and fruitfly, the assessment of the X/A ratio is believed to set the functional state of genes that coordinately govern both dosage compensation and sexual phenotype. *A priori*, these two processes need not be controlled together, and

FIG. 4 Probable functions and regulation of the *Sxl* gene (schematic). It is likely that the early embryonic transcription of *Sxl* 'primes' the gene in XX embryos, so that it remains active thereafter (see text). The gene is known to affect dosage compensation and gametogenesis, as well as somatic sexual phenotype, but the details of these interactions are at present unknown.



in mammals their control is indeed independent, sex being determined by the presence or absence of a Y chromosome, and dosage compensation by the number of X chromosomes. In *D. melanogaster* and *C. elegans*, however, control is unified—in the fly by the *Sxl* gene²³, and in the nematode by the *xol-1*, *sdc-1*, and *sdc-2* genes²⁴⁻²⁶.

A combination of genetic and molecular methods^{23,27-29} has led to a fairly clear outline of the function and regulation of *Sxl* (Fig. 4). This gene responds to the dosage of *sis* elements, acting in early embryogenesis, and becomes set in one of two modes, either active (in females) or inactive (in males). The early phase of *Sxl* control also involves other gene activities such as *da* (*daughterless*) and *snf* (*sans-fille*)^{4,29-31}. Once established in the female mode, *Sxl* acts as a regulator for four processes: the three branches of sexual differentiation, and its own activity. In the male mode, it has no functions.

The transcription patterns of *Sxl*, and the predicted sequences of its products^{27,28}, make it very likely that both its regulatory and autoregulatory functions are achieved by differential control of splicing. In males, all post-embryonic transcripts include an exon that contains an in-frame stop codon, so the resulting truncated protein products are presumably non-functional. In females, an alternative splicing pattern removes this stop codon and full-length proteins can be generated. The predicted products contain a sequence motif found in many RNA-binding proteins²⁷. Therefore, these products may act directly to promote the productive splicing of *Sxl* transcripts, leading to positive autoregulation, and also to control the splicing of target genes such as *tra* (see below).

The initial establishment phase of *Sxl* control may well be achieved by female-specific embryonic transcription²⁸ from a different promoter, activated by *sis* factors. If the splicing of early transcripts avoids the male-specific stop codon, active protein can be produced. Later, identical primary transcripts from a constitutive promoter would be produced in both females and males, but only the transcripts in females would be productively spliced, as a result of the antecedent presence of active *Sxl* protein. Consequently a positive feedback loop is set up, and the *Sxl* gene will remain active in females even though the initiating signal has disappeared.

There are several reasons why such autoregulation may be necessary²³. One is that if dosage compensation is able to affect numerator activity, and if *Sxl* were able to respond continuously to the X/A ratio, then the system would be unstable. An XY embryo, with *Sxl* inactive, would adopt the male mode of increased X chromosome transcription; this would lead to an apparent increase in X/A ratio, which in turn would activate *Sxl*, thereby leading to oscillations in gene activity and no clear difference between the sexes. Autoregulation escapes this problem. A second reason is that a bistable switch of this type may increase the apparent accuracy of the reading of the X/A ratio. Embryonic cells that read the ratio wrongly and adopt an inappropriate *Sxl* state will die as a result of incorrect dosage compensation: consequently mistakes are edited out, and the

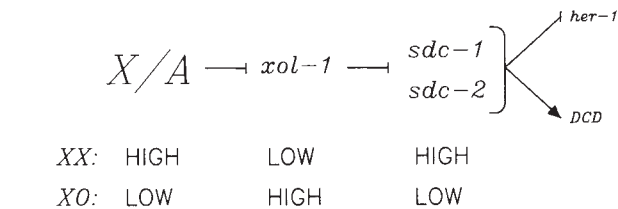


FIG. 5 Genetic model for *sdc* function and control (after ref. 24). The *xol-1* gene is named after the loss-of-function mutant phenotype (*XO lethal*), and the *sdc* genes are named after their suspected functions (*sex and dosage compensation*). Mutations in these two classes of gene have essentially opposite effects (Table 2), and *sdc* mutations are epistatic to *xol-1* mutations, leading to the proposed model. Barred arrows signify negative interactions (down-regulation).

TABLE 1 Comparison of systems

	<i>D. melanogaster</i>	<i>C. elegans</i>
a. Similarities		
X/A ratio as basic signal	XX and X(Y) sexes	XX and XO sexes
Multiple numerator elements on X	<i>sis-a, -b, ...</i>	octamers?
Master regulators of sex and compensation	<i>Sxl</i>	<i>xol, sdc</i>
Dosage compensation genes	<i>MSL</i> set	<i>DCD</i> set
Regulatory cascade for somatic sex	<i>tra, ix, dsx</i>	<i>her, fem, tra</i>
Germline specific controls	<i>ovo?</i>	<i>fog</i>
b. Differences		
Sexes	Female, male	Hermaphrodite, male
Compensation mechanism	Positive	Negative?
Soma and germ line	Different genes	Same genes
Cascade interactions	Positive	Negative
Interaction mechanisms	Alternative splicing	Transcriptional and post-transcriptional

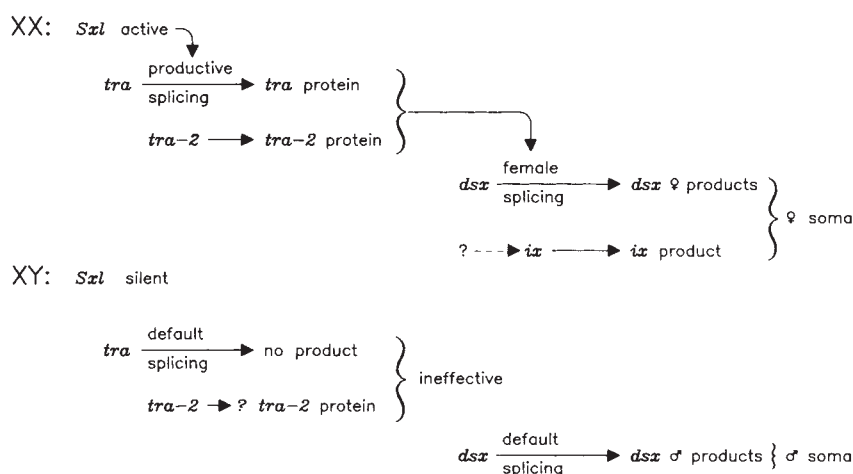
fly embryo has enough cells to compensate for the loss of a few. This type of editing is not an option available to the nematode embryo, which undergoes exact cell lineages and cannot tolerate random cell death; consequently the X/A ratio may have to be read by a more accurate mechanism in the nematode than in the fly.

In addition to regulating its own splicing, *Sxl* also regulates the splicing of the *tra* (*transformer*) gene, which is the first in the cascade of sex-determining genes that act specifically in the *D. melanogaster* soma^{27,32}. As with *Sxl* itself, in the presence of an active *Sxl* gene the *tra* gene transcript is spliced productively, leading to an active feminizing product. In the absence of *Sxl* activity, *tra* transcripts undergo default splicing which results in messages with no long open reading frame.

The other targets of *Sxl* activity, in the dosage compensation and germ-line branches, are at present unknown. Several genes required specifically for dosage compensation have been identified, and are discussed below, but it is not clear which, if any, is subject to splicing (or other) control by *Sxl* products. In contrast to *tra*, which is under positive control by *Sxl*, the dosage compensation genes are likely to be under negative control, because they are required in male flies and not required in female flies. It could be that *Sxl* protein modifies the splicing of one or more of the dosage compensation gene transcripts from a productive into a non-productive pattern, opposite to the effect on *tra*. In the germ line, even less is known about how *Sxl* achieves its effects on gametogenesis.

C. elegans has at least one gene that bears a close functional similarity to *Sxl*. This gene, *sdc-2*, is sex-linked and essential for activating both female sexual differentiation and the female mode of dosage compensation (treating hermaphrodite and female as equivalent for the time being)²⁶. Whether its activity alone is sufficient to activate these processes is uncertain, since there is another gene in this class²⁵, *sdc-1*, but *sdc-1* mutations have less extreme effects than *sdc-2* mutations. Furthermore, a third gene²⁴, *xol-1*, is required for correct expression of *sdc-1* and *sdc-2*. According to the model shown in Fig. 5, *xol-1* acts as an upstream negative regulator of the *sdc* genes in XO animals. However, *xol-1* mutations affect both XX and XO animals²⁴, so the regulation of *xol-1* itself may not be simple. Conceivably, *xol-1* has the same kind of role in the nematode as *da* has in the fly: it might be required for the correct setting

FIG. 6 The somatic cascade inferred for *D. melanogaster*. In XX flies, *Sxl* promotes the productive splicing of *tra* transcripts. The *tra* gene product, together with that of *tra-2*, promotes the female-specific splicing of *dsx* transcripts. In XY flies, both *tra* and *dsx* adopt a default mode of splicing.



of the *sdc* genes, without itself acting as a true switch gene. The two systems would then have greater similarity. If *xol-1* does act as a switch gene, then the X/A ratio acts negatively on *xol-1*, rather than positively on *sdc-1* and *sdc-2*.

There is evidence that interfering with dosage compensation in *C. elegans* can affect the reading of the X/A ratio (so that, for example, mutations in *DCD* genes, which lead to increased X transcript levels, transform 2X; 3A animals from male into hermaphrodite^{33,34}). Therefore, it seems likely that the *sdc* genes would be vulnerable to the potential oscillation outlined above. It would not be surprising if the *sdc* genes were also under autoregulatory control analogous to that operating on *Sxl*.

It is notable that all these master regulators that respond most directly to the X/A ratio are themselves located on the X chromosome, in both fruitfly and nematode. In contrast, most of the other regulatory genes are autosomal. Conceivably, their location on the X chromosome is of ancestral significance. Originally, the *Sxl* and *sdc* genes may have acted as combined numerators and regulators, in which case their location on the X would be essential. Subsequent evolution might have co-opted other sex-linked genes as additional numerators, until eventually the X/A counting mechanism became separate from the regulatory genes themselves.

In summary, a single master regulatory gene, under autoregulatory control, has been identified in *D. melanogaster*. In *C. elegans*, the interactions at this level appear at present to be more complex, but still may involve only two or three genes.

Positive and negative dosage compensation

In both fruitfly and nematode, about 20% of the known genes lie on the X chromosome. The resulting large difference in gene dosage between one sex (XX) and the other (XO or XY) must be compensated for to produce a viable organism. In both species, there is good evidence that failure to achieve correct dosage compensation is lethal. Death can result either from inadequate expression from the single X in one sex, or from overexpression from both X chromosomes in the other. Both species use the mechanism of equalizing transcript levels, probably by affecting transcription. Regulation of transcription has been demonstrated directly in *D. melanogaster* by measuring tritiated uridine incorporation on polytene X chromosomes (for review see ref. 2), but it remains possible that turnover, rather than transcription, is governing transcript levels in *C. elegans*³⁵. In both species, both X chromosomes remain active in the XX sex. Mammals, in contrast, achieve compensation by inactivating one of the X chromosomes in the XX sex.

Experiments on transgenic *D. melanogaster*³⁶ indicate that dosage compensation acts by a two-fold increase of sex-linked gene transcription in male flies, relative to the basal level for an autosomal gene. This is achieved by *cis*-acting elements which

act as two-fold enhancers, apparently capable of placing any nearby gene under dosage compensation control. Some progress has been made in defining these enhancers, for example at the *white* locus³⁷. Several *trans*-acting factors required for increased transcription have also been identified, in the form of the genes *mle* (*maleless*) and *msl-1*, *-2*, *-3* (*male-specific lethal*). Mutations in any of these genes (the *MSL* set) have no effect on XX flies, but lead to the death of XY flies as a result of inadequate X chromosome transcription. Sexual phenotype does not affect this lethality (so that somatically male *tra*⁻ XX flies are not killed by *MSL* mutations). Conversely, the *MSL* mutations do not have any effects on sexual phenotype, only on viability. Therefore, these genes appear to be involved only in the dosage compensation branch of the sex determination network.

It is likely that additional dosage compensation genes remain to be discovered in *D. melanogaster*, because XY flies that carry a feminizing allele of *Sxl*, *Sxl*^{M1}, die earlier than do *msl*⁻ XY flies, although in both cases death appears to be due to inadequate X chromosome transcription. Effects on transcription by the *MSL* set have only been demonstrated relatively late, but it has been shown that dosage compensation is already operative by the blastoderm stage³⁸.

In nematodes, experiments on transgenic animals have not yet reached a point where one can decide whether compensation occurs by a positive mechanism acting in XO animals, or by a negative mechanism acting in XX animals, or both. However, a set of at least four autosomal genes affecting dosage compensation have been identified, which have opposite properties to the *MSL* set. Mutations in these genes are deleterious or lethal to XX animals but not to XO animals, independent of phenotypic sex³⁴. These four genes are known as the 'dosage compensation dumpies' (*DCD* set: *dpy-21*, *-26*, *-27*, *-28*) because the sub-lethal phenotype in XX mutants is a characteristic dumpy body shape, seen also in 3X; 2A individuals. Furthermore, several of these mutations have been shown to cause raised levels of sex-linked transcripts in XX animals, but not in XO animals^{34,38}. One cannot exclude the possibility that these genes act as negative regulators for nematode genes of the *msl* type, but extensive searches for mutations of the *msl* type have failed to yield any good candidates⁸. Therefore, it may be that compensation in the nematode works in an opposite direction from that in the fruitfly: specific regulatory genes are activated in XX animals that reduce sex-linked transcript levels twofold, thereby equalizing XX and XO.

In both organisms, it will be extremely interesting to learn more about the molecular mechanisms involved in dosage compensation. These mechanisms achieve a twofold increase or decrease in the transcript level for a large number of functionally unrelated genes (at least 500 genes in *C. elegans* and 1,000 in

D. melanogaster) and must therefore be able to work correctly in a great variety of regulatory contexts.

Somatic sexual phenotype

In both nematode and fruitfly, somatic sexual phenotype is controlled by a small number of autosomal regulatory genes, organized in an ordered cascade of interactions. These genes do not affect dosage compensation, and several lines of evidence confirm that they act downstream of the sex-linked master regulators. In *D. melanogaster*, there are four major genes in this category³⁹: *tra*, *tra-2*, *ix* and *dsx*. In *C. elegans* there are seven⁷: *her-1*, three *fem* genes and three *tra* genes. An important difference between the two systems is that the fruitfly somatic sex-determination genes have no determinative effect on the germ line, but in the nematode the same set of genes functions in determining sexual phenotype in both soma and germ line. However, the functions and interactions of the genes are rather different in soma and germ line, so it is simpler to consider the two branches independently.

In *D. melanogaster*, genetic data^{39,40} indicate that *Sxl* controls *tra* and/or *tra-2*, and one or both of these genes controls *dsx*, and possibly *ix* as well. Molecular evidence has confirmed and extended this model (Fig. 6). As described above, the active *Sxl* gene product in female flies directs the productive splicing of *tra* transcripts, resulting in active *tra* protein in XX flies but not in XY flies. A transgenic XY fly carrying a cDNA for the female-specific *tra* transcript is completely feminized in the soma, demonstrating that this activity alone can direct female somatic development⁴¹. The second gene, *tra-2*, is not obviously regulated in the soma, since identical transcripts are present in both male and female flies^{42,43}. However, the predicted structure of the protein product of *tra-2* contains the same RNA-binding motif found in *Sxl*, and *tra* is related structurally to other fruitfly genes implicated in splicing. So it is possible that these two products act in conjunction to regulate the splicing of a target gene. In the absence of *tra* product, the *tra-2* protein would be present but ineffective.

The obvious candidate for a target gene is *dsx* (*doublesex*), which, like *Sxl* and *tra*, shows sex-specific differences in splicing⁴⁴. In contrast to the other two genes, however, both the male and female products of *dsx* appear to be functional proteins. This is consistent with the genetic analysis of *dsx*, which has functions in both female and male development³⁹. Null mutations lead to intersexual development in the soma of both

TABLE 2 Examples of mutant phenotypes (effects on soma and viability)

	XX phenotype	XY or XO phenotype
<i>Drosophila melanogaster</i>		
Wild type	Female	Male
<i>Sxl(lf)</i>	Masculinized, lethal	Male
<i>Sxl(gf)</i>	Female	Feminized, lethal
<i>msl-1(lf)</i>	Female	Male, lethal
<i>tra(lf)</i>	Male	Male
<i>tra(gf)</i>	Female	Female
<i>Caenorhabditis elegans</i>		
Wild type	Female	Male
<i>sdv-2(lf)</i>	Masculinized, lethal	Male
<i>xol-1(lf)</i>	Female	Feminized, lethal
<i>dpy-26(lf)</i>	Female, lethal	Male
<i>tra-1(lf)</i>	Male	Male
<i>tra-1(gf)</i>	Female	Female

This table summarizes a few of the more significant mutants in each system. Mutant phenotypes are marked in bold. Viability effects are due to improper dosage compensation. Germ-line phenotypes for some of these mutations are more complicated, so have been omitted from this table (see text). The wild-type *C. elegans* hermaphrodite appears to be completely female with respect to somatic development, so the soma is classified as female. For convenience of comparison, alleles have been abbreviated *lf*, for loss-of-function, or *gf*, for gain-of-function. The *lf* alleles are usually recessive, the *gf* alleles usually dominant. All the genes discussed in this article have been defined by at least one *lf* allele, and in many cases candidate null alleles have been identified. For three of the *D. melanogaster* genes and four of the *C. elegans* genes, both *lf* and *gf* alleles have been generated.

XX and XY flies. The gene therefore seems to have a dual role: in XX flies it represses male development, permitting female development to occur; in XY flies it represses female development and permits male development. The fourth gene, *ix* (*intersex*), is required for the female, but not the male, functions of *dsx*. As yet, there is no information as to whether *ix* is itself subject to regulation.

In *C. elegans*, the cascade involves more genes and more steps. This greater complexity may possibly reflect the dual roles of these genes in both soma and germ line, which may require more elaborate interactions. The genetic data indicate that there are at least four steps in the cascade, involving negative regulatory interactions at each step, as set out in Fig. 7. The activity of the last gene in the cascade, *tra-1* (*transformer-1*), is both necessary and sufficient to direct all aspects of female somatic development⁴⁵. Thus, a dominant gain-of-function mutation of *tra-1* leads to female development of an XO animal, irrespective of the state of the nine upstream regulatory genes. Conversely, absence or low activity of *tra-1* is both necessary and sufficient for male somatic development, and null mutations of *tra-1* lead to complete somatic masculinization (Table 2).

Five of the seven genes in the cascade have now been cloned and partly or wholly sequenced (*her-1*, C. Trent & W. B. Wood, personal communication; *tra-2*, P. Okkema, P. Kuwabara & J. Kimble, personal communication; *fem-3*, ref. 46; *fem-1*, ref. 47; *tra-1*, D. Zarkower & J.H., unpublished data). Thus far, none of the genes bears any similarity to any of the *D. melanogaster* sex determination genes, nor (with one exception) to other eukaryotic regulatory genes. The exception is the *fem-1* gene⁴⁷, which encodes a predicted protein of 67K (*M_r* 67,000); about one third of the protein is composed of six repeats of a motif first described in certain 'start'-associated cell-cycle genes in fission and budding yeasts. The same six-fold repeat appears in the cytoplasmic domains of proteins that govern cell-cell interactions in *D. melanogaster* (*Notch*) and *C. elegans* (*lin-12*, *glp-1*), but its significance is presently unknown⁴⁸⁻⁵¹.

Transcriptional analysis of the five genes has so far provided no evidence for control at the level of alternative splicing. The *her-1* gene seems to be transcriptionally regulated, in exactly the manner predicted by the genetic model. Transcripts are

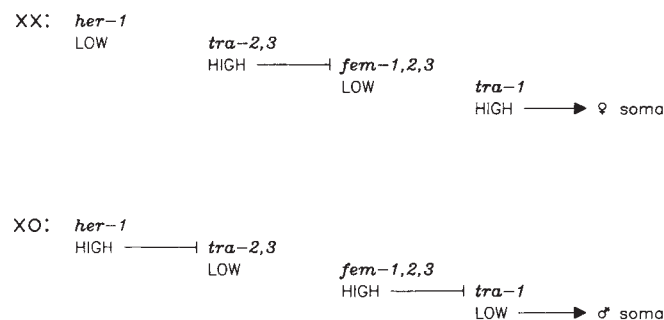
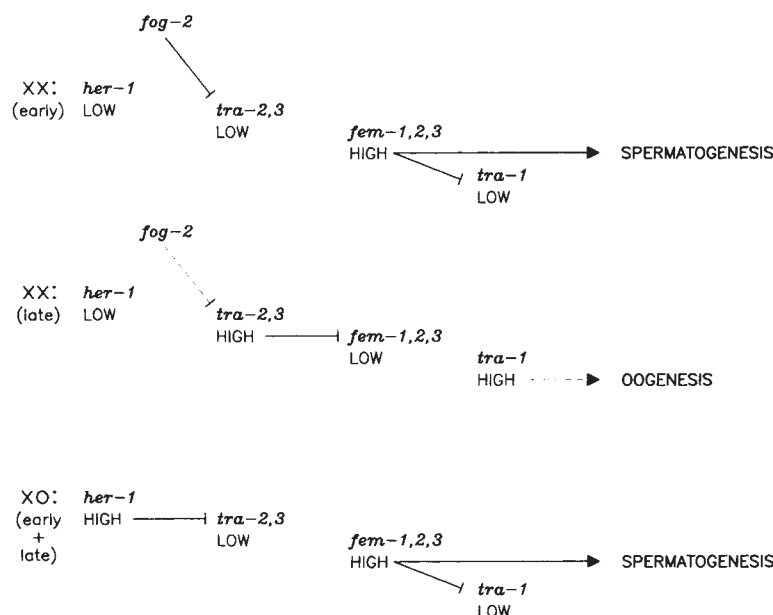


FIG. 7 The somatic cascade inferred for *C. elegans* (slightly simplified). Both *her* (*hermaphroditization*) and *fem* (*feminization*) are named after the loss-of-function mutant phenotype. Their wild type function is opposite, i.e. masculinization. The wild type function of the nematode *tra* genes is to feminize, like the *tra* genes in *D. melanogaster*. Interactions between successive steps in the cascade are negative, symbolized by barred arrows (their molecular nature is at present largely unknown). Consequently the seven genes adopt opposite activity states in the two sexes. In XX animals, *her-1* activity is low (down-regulation by *sdv* genes, as in Fig. 5), so the *tra-2* and *tra-3* genes are allowed to be active. This leads to low activity of the *fem* genes, and consequently to high activity of *tra-1*. In XO animals, *her-1* is at high activity and this leads to low activity of *tra-2* and *tra-3*, high activity of the *fem* genes, and low activity of *tra-1*. The activity of the last gene in the cascade, *tra-1*, determines somatic sexual phenotype.

FIG. 8 The germ-line cascade in *C. elegans*. Several slightly different models for germ-line modulation are possible; the one shown is the simplest, but it is also possible^{10,53,60} that modulation acts directly on the *fem* genes, as well as through *tra-2*. The evidence for this kind of model is first that regulation at the *her-1* level of the cascade affects a choice between male and hermaphrodite in the germ line, but regulation at the *fem* level affects a male/female choice. Second, both loss-of-function mutations of *fog-2* and certain gain-of-function mutations of *tra-2* selectively eliminate spermatogenesis from XX animals but do not affect XO animals (thereby converting the sexual system into a conventional fertile female/fertile male arrangement)^{59,60}. The interpretation is that these *tra-2(gf)* mutations render the *tra-2* gene insensitive to the germ-line control by *fog-2*, but do not affect control by *her-1*. Additional genes are also involved in germ-line sex determination^{10,60}, for example the gene *fog-1*. The role of *tra-1* in the germ line is problematic^{45,58}, because *tra-1* appears to be important, but not essential, for normal gametogenesis in both hermaphrodites and males.



present at much higher level in XO males or in XX animals that have been masculinized by an *sd-2* mutation, than in XX hermaphrodites or XX animals that have been masculinized by a *tra-1* mutation (C. Trent and W. B. Wood, personal communication). Of the other genes, neither *fem-1* nor *fem-3* seems to have a male-specific transcript^{46,47}, although there is evidence for a higher level of *fem-3* transcripts in male embryos, again as predicted by the genetic model. Transcription from *tra-2* (P. Okkema and J. Kimble, personal communication) and *tra-1* (D. Zarkower and J.H., unpublished data) appears to be complex. Genetic arguments have been advanced that some of the regulation in the pathway must occur at a post-transcriptional level^{52,53}.

Both *dsx* in *D. melanogaster* and *tra-1* in *C. elegans* act as global regulators, affecting a large number of different developmental processes in different ways. Mosaic analysis of *D. melanogaster*³⁹ and *C. elegans* (C. Hunter and W. B. Wood, personal communication) has shown that these genes act in a cell-autonomous manner, so they are unlikely to encode hormones. How these two genes carry out their multitude of functions is beyond the scope of this review, but is the subject of much research.

In summary, the *D. melanogaster* somatic cascade involves a series of positive regulatory interactions, associated with alternative splicing patterns, whereas the *C. elegans* somatic cascade involves negative regulatory interactions and there is no evidence for regulation at the level of splicing, but some evidence for regulation at other levels. At present, the molecular similarity between the two systems seems to be minimal.

Fruitfly germ-line sexual phenotype

In *D. melanogaster*, the *Sxl* gene is necessary for oogenesis but not for spermatogenesis⁵⁴, whereas the somatic sex determination genes have no determinative role in the germ line. This implies the existence of separate regulatory events that determine whether sperm or oocytes are made. However, germ-line sexual phenotype is not wholly independent of somatic phenotype, because XX germ cells can be diverted into the male (spermatogenic) pathway if transplanted into a male somatic environment⁵⁵. This masculinization does not occur if the transplanted germ cells carry a dominant feminizing mutation of *Sxl*, providing further evidence for the involvement of *Sxl*. These effects of the soma on the germ line are reminiscent of those seen in mammals⁵⁶.

Many genes specifically required for oogenesis or spermatogenesis have been studied in *D. melanogaster*, but it is not clear if any of them actually governs the choice between the

two gamete types. For example, *ovo* mutations have no effect on spermatogenesis and lead to abortive oogenesis, but there is little evidence that the XX germ line in *ovo* mutants is being diverted into a male pathway⁵⁷. Consequently, there are as yet no candidates for the germ-line targets of *Sxl* activity.

The *tra-2* gene of *D. melanogaster* is unusual in that it has a role in the germ line as well as the soma^{39,42,43}. This role is male specific, in contrast to its female-specific somatic function. XY *tra-2* male flies are sterile, because spermatogenesis is arrested, but there is no evidence that sexual phenotype is being affected. As described above, the state of *Sxl* probably does not affect whether *tra-2* products are made, so the spermatogenic function of *tra-2* can be ascribed to a function that it can carry out in the absence of *tra* product. In the soma, its activity is dependent on the presence of *tra* product.

Nematode germ-line sexual phenotype

Considerably more is known about how germ-line sex is specified in the nematode than in the fruitfly. As described at the beginning of this article, the germ line of the XO sex is male, but the germ line of the XX sex is mixed: initially male and then female, thereby enabling reproduction by self-fertilization. The male phase of gametogenesis occurs inside a female soma, so during hermaphrodite spermatogenesis, germ line and soma are undergoing opposite sexual differentiation.

The same seven genes that control somatic sex in *C. elegans* also control germline sex, but there are several important differences in their functions and interactions. First, the *fem* genes are likely to have a dual role in the germ line, acting to masculinize both directly, by activating spermatogenesis, and indirectly, by down-regulating *tra-1* (as they do in the soma)⁵². In the absence of any of the *fem* gene activities, only oocytes are made, whatever the state of *tra-1*. Second, the role of *tra-1* is complicated^{45,58}. Dominant gain-of-function mutations feminize both germ line and soma, consistent with the general role of *tra-1* as a feminizing gene, but loss-of-function mutations have paradoxical effects on the germ line, leading to reduced gametogenesis and sometimes to oogenesis in both XX and XO individuals, although these are somatically male. Some level of *tra-1* activity therefore appears to be necessary for correct gametogenesis in both sexes.

The third difference is a modulation specific to the germ line that permits the initial phase of spermatogenesis in XX animals^{59,60}. The simplest model for this modulation is set out in Fig. 8. The model postulates that a specific germ-line activity, due to the gene *fog-2* (*feminization of germ line*) and perhaps

others, leads to a transient down-regulation of the *tra-2* gene in the germ line. This allows the *fem* genes to be active and promote spermatogenesis.

This model implies that *tra-2* activity is under general negative control in XO animals (presumably by *her-1*) and also under germ-line specific negative control in XX animals (presumably by *fog-2*). Several gain-of-function mutations of *tra-2* with altered regulatory properties have now been sequenced (P. Okkema & J. E. Kimble, personal communication), which should shed light on the molecular mechanisms involved in these two kinds of regulation.

In summary, relatively little is known about how sex is governed in the *D. melanogaster* germ line, but the somatic control genes are not involved. In *C. elegans*, the same genes are used in both soma and germ line, but the interactions are not the same in the two branches.

Discussion

Research on sex determination and sexual differentiation in these two species impinges on many general biological problems: measurement of thresholds, autoregulation, fine tuning of gene expression, tissue-specific gene interactions and the coordination of gene activities in different tissues. The experimental advantages have been particularly favourable, and the networks of regulatory genes sufficiently simple (perhaps no more than 20 genes involved in either species) that one can hope for a reasonably complete molecular description in each case.

The comparison of the two systems provides at present the best opportunity for detailed comparison of the same developmental process in two widely diverged organisms. It should be clear that there is similarity in overall organization but no resemblance in terms of molecular biology.

Does this mean that developmental geneticists should despair of extrapolating from one organism to another? Probably not, for three reasons. First, it is already known that the components of some developmental mechanisms are strongly conserved at the molecular level. The *Notch* (*D. melanogaster*) and *lin-12/glp-1* (*C. elegans*) genes, which govern cell-cell interactions in both species, provide a fine example⁴⁸⁻⁵¹; so do several aspects of homoeobox gene function and organization, which may be conserved between flies, nematodes and mammals.

The second reason is that sex determination mechanisms seem to change much more rapidly in evolution than do other components of development, so the lack of molecular conservation is not surprising⁶¹. Nevertheless, the information gained from these detailed studies of *D. melanogaster* and *C. elegans* should allow us to work outwards, examining related species with different sex-determination systems, and thereby to explore the evolutionary modification of regulatory networks. For example, it should be possible to use the cloned genes from *C. elegans* to isolate and analyse corresponding genes from closely related species such as *C. remanei*, which has a male/female sex-determination system.

Finally, one can make useful extrapolations at an organizational level, for example in predicting that some kind of autoregulatory step will be found in *C. elegans*, and that genes specifically controlling germ-cell identity will be found in *D. melanogaster*. In turn, one can hope to identify general principles that can be applied widely in developmental and evolutionary studies, even if completely dissimilar molecules are used to implement these principles in different animal groups. □

Jonathan Hodgkin is at the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. Address until October 1990: Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, USA.

1. Baker, B. S. & Belote, J. M. *Rev. Genet.* **17**, 345-393 (1983).
2. Lucchesi, J. C. & Manning, J. E. *Adv. Genet.* **24**, 371-429 (1987).
3. Baker, B. S. *Nature* **340**, 521-524 (1989).
4. Cline, T. W. *Cell* **59**, 231-234 (1989).
5. Wolfner, M. *Trends Genet.* **4**, 333-337 (1988).
6. Wood, W. B. *et al.* (eds) *The Nematode Caenorhabditis elegans* (Cold Spring Harbor, New York, 1988).
7. Hodgkin, J. A. *Rev. Genet.* **21**, 133-154 (1987).
8. Villeneuve, A. M. & Meyer, B. J. *Adv. Genet.* **27**, 117-188 (1990).
9. Meyer, B. J. *Trends Genet.* **4**, 337-342 (1987).
10. Kimble, J. *Phil. Trans. R. Soc. Lond. B* **322**, 11-18 (1988).
11. White, M. J. D. *Animal Cytology and Evolution* (Cambridge University Press, Cambridge, 1973).
12. Sudhaus, W. *Zool. Jb. Syst. Bd.* **101**, 417-465 (1974).
13. Bridges, C. B. *Science* **54**, 252-254 (1921).
14. Madl, J. A. & Herman, R. K. *Genetics* **93**, 393-402 (1979).
15. Dobzhansky, T. & Schultze, J. *Proc. natn. Acad. Sci. U.S.A.* **17**, 513-518 (1931).
16. Cline, T. W. *Genetics* **119**, 829-862 (1988).
17. Meneely, P. M. & Nordstrom, K. D. *Genetics* **119**, 365-375 (1988).
18. Torres, M. & Sanchez, L. *EMBO J.* **8**, 3079-3086 (1989).
19. Alonso, M. C. & Cabrera, C. V. *EMBO J.* **7**, 2585-2591 (1988).
20. Murre, C., McCaw, P. S. & Baltimore, D. *Cell* **56**, 777-783 (1988).
21. Romani, S., Campuzano, S. & Modolelli, J. *EMBO J.* **6**, 2085-2092 (1987).
22. McCoubrey, W. K., Nordstrom, K. D. & Meneely, P. M. *Science* **242**, 1146-1151 (1988).
23. Cline, T. W. *Genetics* **107**, 231-277 (1984).
24. Miller, L. M., Plenefisch, J. D., Casson, L. P. & Meyer, B. J. *Cell* **55**, 167-183 (1988).
25. Villeneuve, A. M. & Meyer, B. J. *Cell* **48**, 25-37 (1987).
26. Nussbaum, C. & Meyer, B. J. *Genetics* **122**, 579-593 (1989).
27. Bell, L. R., Maine, E. M., Schedl, P. & Cline, T. W. *Cell* **55**, 1037-1046 (1988).
28. Salz, H. *et al.* *Genes Dev.* **3**, 708-719 (1989).
29. Cline, T. W. *Genetics* **113**, 641-663 (1986).
30. Oliver, B., Perrimon, N. & Mahowald, A. P. *Genetics* **120**, 159-171 (1988).
31. Steinmann-Zwicky, M. *EMBO J.* **7**, 3889-3898 (1988).
32. Nagoshi, R. N. *et al.* *Cell* **53**, 229-236 (1988).
33. Hodgkin, J. *Development* **101** Suppl., 5-16 (1987).
34. Plenefisch, J. D., DeLong, L. & Meyer, B. J. *Genetics* **121**, 57-76 (1989).
35. Meyer, B. J. & Casson, L. P. *Cell* **47**, 871-881 (1986).
36. Spradling, A. C. & Rubin, G. M. *Cell* **34**, 47-57 (1983).
37. Hazelrigg, T. *Trends Genet.* **3**, 43-47 (1987).
38. Gergen, P. *Genetics* **117**, 477-485 (1987).
39. Baker, B. S. & Ridge, K. *Genetics* **94**, 383-423 (1980).
40. Cline, T. W. *Dev. Biol.* **72**, 266-275 (1979).
41. McKeown, M., Belote, J. M. & Boggs, R. T. *Cell* **53**, 887-895 (1988).
42. Amrein, H., Gorman, M. & Nothiger, R. *Cell* **55**, 1025-1035 (1988).
43. Gorski, T., Edstrom, J. E. & Baker, B. S. *Cell* **56**, 1011-1018 (1989).
44. Burtis, K. C. & Baker, B. S. *Cell* **56**, 997-1009 (1989).
45. Hodgkin, J. *Genes Dev.* **1**, 731-745 (1987).
46. Rosenquist, T. A. & Kimble, J. E. *Genes Dev.* **2**, 606-616 (1988).
47. Spence, A. M., Coulson, A. & Hodgkin, J. *Cell* (in the press).
48. Wharton, K. A. *et al.* *Cell* **43**, 567-581 (1985).
49. Kidd, S., Kelley, M. R. & Young, M. W. *Molec. cell. Biol.* **6**, 3094-3098 (1986).
50. Yochem, J. & Greenwald, I. *Cell* **58**, 553-563 (1989).
51. Austin, J. & Kimble, J. *Cell* **51**, 589-599 (1987).
52. Hodgkin, J. *Genetics* **114**, 15-52 (1986).
53. Barton, M. K., Schedl, T. B. & Kimble, J. *Genetics* **115**, 107-119 (1987).
54. Schupbach, T. *Genetics* **109**, 529-548 (1985).
55. Steinmann-Zwicky, M., Schmid, H. & Nothiger, R. *Cell* **57**, 157-166 (1989).
56. McLaren, A. *Phil. Trans. R. Soc. Lond. B* **322**, 3-9 (1988).
57. Oliver, P., Perrimon, N. & Mahowald, A. *Genes Dev.* **1**, 913-923 (1987).
58. Schedl, T., Graham, P. L., Barton, M. K. & Kimble, J. E. *Genetics* **123**, 755-769 (1989).
59. Doniach, T. *Genetics* **114**, 53-76 (1986).
60. Schedl, T. & Kimble, J. E. *Genetics* **119**, 43-61 (1988).
61. Bull, J. J. *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings, Menlo Park, California, 1984).

ACKNOWLEDGEMENTS. I thank those mentioned in the text for permission to cite unpublished results, and J. Becker for drawing the figures. V. Bardwell, J. Kimble, B. J. Meyer, A. M. Spence and D. Zarkower provided especially valuable criticisms of this review. I acknowledge support from the Medical Research Council, the Parke Davis Exchange Fellowship Fund, and the University of Wisconsin (Brittingham Visiting Professorship Program).

Stratospheric ozone depletion and future levels of atmospheric chlorine and bromine

Michael J. Prather & Robert T. Watson

NASA Headquarters, Code EEU, Washington, DC 20546, USA

The rise in atmospheric chlorine levels caused by the emission of chlorofluorocarbons and other halocarbons is thought to be the main cause of the appearance of the Antarctic ozone 'hole' in the late 1970s, and the more modest ozone depletion observed over parts of the Northern Hemisphere. Atmospheric bromine, also associated with halocarbon emissions, is believed to contribute to ozone depletion. Over the next decade, further increases in these compounds are inevitable. Model calculations show that by the end of the next century, atmospheric chlorine and bromine levels may return to those prevalent before the onset of the ozone hole, but only if more stringent regulations are applied to halocarbon production than those currently proposed.

HERE we look at projected future levels of stratospheric chlorine for those compounds currently regulated under the 1987 Montreal Protocol¹ (CFC-11, CFC1₃; CFC-12, CF₂Cl₂; CFC-113, CF₂ClCFC1₂; CFC-114, CF₂ClCF₂Cl; CFC-115, CF₃CF₂Cl), for other gases not specified in the current Montreal Protocol (carbon tetrachloride, CCl₄; methyl chloroform, CH₃CCl₃; HCFC-22, CHF₂Cl), and for possible new hydrochlorofluorocarbons (HCFCs) suggested as substitutes for the CFCs. A subset of these scenarios is also examined to study the effects of possible future increases in atmospheric bromine, resulting from emissions of halons as listed under the 1987 Montreal Protocol (CF₃Br, CF₂ClBr, CF₂BrCF₂Br) assuming constant levels of CH₃Br. All these halocarbons accumulate in the troposphere, are then readily transported into the stratosphere, and along with CH₃Cl, constitute the source of reactive chlorine and bromine species that destroy O₃. We focus on the bulk atmospheric abundances of these halocarbons as a direct measure of the concentration of chemically active chlorine and bromine species in the stratosphere, and hence, of the potential damage to stratospheric ozone.

The total stratospheric abundance of chlorine compounds today is about 3 p.p.b. (parts per 10⁹ molecules of air) and consists predominantly of the industrial halocarbons and their photochemical byproducts (for example, HCl and ClO). Reactions involving chlorine monoxide (ClO) and atomic chlorine (Cl) are able to catalyse the destruction of stratospheric ozone on a global scale, and the ClO dimer plays a key part in the formation of the Antarctic ozone hole². Bromine in the stratosphere, present at concentrations of ~0.02 p.p.b. also contributes globally to ozone loss^{3,4}. Concentrations of BrO and ClO are greatly enhanced in the winter stratosphere over the Antarctic and Arctic. During the austral spring of 1987, bromine was estimated to be responsible for between 10 and 30% of the ozone depletion with chlorine being responsible for the remainder⁵⁻⁷.

Atmospheric chlorine concentrations have increased from 0.6 p.p.b. a century ago to 2 p.p.b. in the late 1970s when the ozone hole was recognized to have first occurred^{8,9} and are now at

more than 3 p.p.b. In the Southern Hemisphere the ozone loss over Antarctica, amounting to about 50% of the ozone column in the spring, coincides with significant declines (as much as 10%) over the mid-latitudes during spring and summer¹⁰. In northern mid-latitudes, we have observed more modest ozone decreases in late winter and early spring, ranging from 3 to 6% in the ozone column over the past two decades¹¹. Recovery of the Antarctic ozone hole as we understand it, would not occur until chlorine levels fall below 2 p.p.b. (henceforth, the 2-p.p.b. date), although recovery might also depend on changes in the Antarctic climate and concentrations of other trace gases such as methane.

The largest ozone depletions in the future are expected to coincide with the peak chlorine loading of the atmosphere. Losses in the middle and upper stratosphere are driven by chemical reactions of chlorine monoxide with atomic oxygen and occur on a global scale. Using only these gas-phase reactions, the ozone depletion predicted as chlorine increases from 3 to 5 p.p.b. is modest, about 1-2% in the tropics and 4-6% at high latitudes¹². In the lower stratosphere, ozone depletion through chlorine- and bromine-catalysed reactions is greatly enhanced by heterogeneous reactions occurring on the surfaces of polar stratospheric cloud particles, and occurs at high latitudes during late winter and early spring. At present, global chemical models are unable to predict how the depth or extent of the ozone hole will change when chlorine increases to 5 p.p.b.. The rapid deepening of the Antarctic ozone hole over the last decade demonstrates the extremely nonlinear response of polar ozone to increasing chlorine levels. We must be prepared for the possibility of new thresholds of accelerated ozone loss in the Arctic as chlorine concentrations reach new record levels.

Chlorine loading and ODPs

The relative contribution of individual source gases to total atmospheric chlorine for the year 1985 is shown below. The known natural sources of chlorine make up only 20% of the total, and the CFCs listed under the Montreal Protocol account for about 50%. The chlorine loading calculated here is the instantaneous, bulk tropospheric mixing ratio (p.p.b.) of chlorine atoms in the form of all the halocarbons listed below.

Industrial sources	
CFC-11	22%
CFC-12	25%
CFC-113	3%
CFC-114	<1%
CFC-115	<1%
HCFC-22	3%
carbon tetrachloride	13%
methyl chloroform	13%
Natural sources	
methyl chloride	20%

This definition is a conservative measure of the amount of stratospheric chlorine that may be active in catalytic ozone destruction.

Ozone depletion potentials (ODPs) have been used to measure the relative contribution of different halocarbons toward destroying ozone^{13,14}. The ODP of a halocarbon is calculated using a chemical model for stratospheric ozone and assuming that the gas is in steady state between its emission and removal

TABLE 1 Halocarbon model

Halocarbon		Lifetime (yr)	Flux (kt yr ⁻¹)	Growth (kt yr ⁻¹ yr ⁻¹)	Concentration (p.p.t.)	Factor (kt p.p.t. ⁻¹)
CFC-11	CFC1 ₃	60	360	14.4	220	23.2
CFC-12	CF ₂ Cl ₂	120	450	18.0	375	20.4
CFC-113	C ₂ F ₃ Cl ₃	90	165	6.6	30	31.6
CFC-114	C ₂ F ₄ Cl ₂	200	15	0.6	5	28.9
CFC-115	C ₂ F ₅ Cl	400	10	0.4	4	26.1
carbon tetrachloride	CCl ₄	50	80	3.2	100	25.9
HCFC-22	CHF ₂ Cl	15	140	5.6	80	14.6
methylchloroform	CH ₃ CCl ₃	6	600	24.0	130	22.4
methylchloride	CH ₃ Cl	1.5			600	
(sub. X)	C _x Cl	15			0	19.3
(sub. Y)	C _x Br	6			0	19.3
halon-1211	CF ₂ BrCl	15	10	0.40	1.5	27.9
halon-1301	CF ₃ Br	110	9	0.36	1.7	25.1
methylbromide	CH ₃ Br	1.5			15	

Lifetime (yr) is the global abundance divided by the annual average loss rate; values are taken from recent reports^{12,14} using the AFEAS review¹⁴ for species destroyed predominantly in the troposphere (CH₃CCl₃ and HCFC-22). Fluxes (kt yr⁻¹ = 1,000 tonnes per year) are based on estimates of production in the year 1985 and total 1,000,000 metric tons for CFCs alone. Growth (kt yr⁻¹ yr⁻¹), the increment to the flux each year from 1986 through 1990, is assumed to be 4% yr⁻¹ for all industrial halocarbons; such increases are conservative estimates of growth in the production of CFC-11 and CFC-12 during this period¹⁷. Concentration is the average tropospheric mixing ratio used to initialize the model and is based on available observations¹¹. Factor (kt p.p.t.⁻¹) relates concentration (p.p.t.) to global content (kt), assuming the compound is well mixed throughout 95% of the atmosphere. Methyl chloride and methyl bromide concentrations are fixed. The concentration (CONC) in YR + 1 is calculated recursively from the equation CONC[YR + 1] = CONC[YR] × DECAY + FLUX[YR + 1] × (1 - DECAY) × LIFETIME/FACTOR where DECAY = exp(-1/LIFETIME).

from the atmosphere. For most CFCs the time taken for atmospheric concentrations to approach within 10% of steady-state conditions is more than 100 years. ODPs are usually defined relative to CFC-11 (dimensionless) rather than in absolute units. ODPs therefore reflect the relative contribution of CFCs towards ozone destruction after nearly constant emissions for a century. Transient ozone depletion potentials have also been used^{12,13}, but they do not clearly reflect the absolute contribution of different halocarbons to atmospheric chlorine concentrations.

Most of the chlorine atoms in CFC-11 are photolysed in the lower stratosphere and thus are available to participate in the catalytic destruction of ozone at all altitudes in the stratosphere. Other CFCs, such as CFC-12 and HCFC-22, are photolysed more slowly, and most of their chlorine atoms are available only in the upper stratosphere or near the winter poles. (Air in the winter polar stratosphere seems to be photochemically aged^{15,16}, having descended from the upper stratosphere.) For such compounds, the model-derived ODPs (relative to CFC-11) will be less than the relative chlorine loading in these scenarios. A warning about the accuracy of ODP calculations must be made: the current assessment models are limited to gas-phase chemistry and predict future ozone depletion predominantly in the middle stratosphere; they do not predict the Antarctic ozone hole or the possibility of similar heterogeneously driven loss in the Arctic lower stratosphere.

The model for halocarbons

The data for halocarbons used in this model for chlorine- and bromine-loading are based directly on scenarios developed for the recent international ozone assessment¹². The UNEP/WMO study focused primarily on scenarios leading to increases in atmospheric chlorine; our study examines approaches to reduce chlorine abundance. Two other differences are (1) the lifetime of HCFC-22 is taken here as 15 yr (ref. 14) rather than 20 yr and (2) we assume that the cutback in CFC emissions will occur in a specified year rather than being phased in over the period from 1996 to 2000.

A simple numerical model is used to calculate the atmospheric abundances of the individual halocarbons, as described in Table 1. Fluxes of the chlorine- and bromine-containing species, except CH₃Cl and CH₃Br, are defined in each scenario, and reductions occur instantly at the end of the year. More detailed atmospheric models for the calculation of halocarbon concentrations are available; however, increasing the complexity of the

model is not justified for this work because first-order uncertainties, such as in the atmospheric lifetimes, overwhelm the differences between such models.

The reference scenario is chosen as an optimistic lower baseline for chlorine loading and assumes control over the atmospheric emissions of all industrial halocarbons as a group (CFCs, CCl₄, HCFC-22, CH₃CCl₃ and the halons). It is initialized in 1985, adopts conservative growth rates through 1990¹⁷, freezes emissions into the atmosphere at the 1990 rate, and cuts all emissions to zero at the end of the year 2000. From then on the atmospheric abundances of chlorine and bromine change solely by the slow decay of gases already in the atmosphere. In scenarios where only a partial cutback is assumed, the fraction of the flux allowed to continue refers to 1985 emissions.

Uncertainties

In these calculations the absolute timing of the atmospheric response may be imprecise because of basic uncertainties in modelling atmospheric composition (for example, CFC lifetimes) as well as simplifying assumptions particular to this model. For example, the stratospheric destruction of halocarbons is the predominant sink for CFCs and will lag emissions by about two years, the time it takes for air to travel from the upper troposphere to the middle stratosphere. In turn, the ozone depletion in response to halocarbons will be similarly delayed. Air over the winter pole may represent a mix of tropospheric halocarbons emitted over the past few years. These effects will be important only when tropospheric concentrations are changing rapidly and have not been accounted for here.

A more important and obviously systematic omission in these calculations is the failure to simulate the delay between halocarbon production and emissions. Depending on use, the halocarbons produced in a given year may not be emitted into the atmosphere until several years later. For example, CFC-11 used in blowing closed-cell insulating foams may take more than ten years to escape into the atmosphere, and CFC-12 used in hermetically sealed refrigeration systems will not be released until the refrigerant coils break, on average 20 years later. This accumulation of CFCs makes it difficult to stop emissions (as in the reference scenario) immediately on cessation of production. On the other hand, some applications will have little delay between production and emission: CFC-113 and CH₃CCl₃ are used primarily as cleaning agents, and their production reflects the amount lost during the cleaning process, presumably to the

TABLE 2 Peak chlorine loading and date to reach 2 p.p.b.

Option	Scenario	Peak chlorine (p.p.b.)	2-p.p.b. date
1	1995 (100% cut)	4.24	2055
	2000 (100% cut, ref.)	4.78	2073
	2005 (100% cut)	5.28	2091
2	2000 (90% cut)	4.78	2175
	2000 (80% cut)	4.78	never
4	2000/2015 (90/10% cut)	4.78	2079
	2000/2015 (80/20% cut)	4.78	2084
5	1995/2000 (50/50% cut)	4.24	2063
7	lifetimes -25%	4.47	2053
	lifetimes +25%	5.01	2095

atmosphere. As of 1987 the total amount or 'bank' of CFCs in products with an average retention time of >2 yr is estimated to be <2 yr of the annual production of all CFCs¹⁷. The net effect of such CFC accumulation may have similar results to a delay in compliance with CFC cuts (see option 4 below).

In spite of these uncertainties in the timing of atmospheric effects, the relative changes in peak chlorine loading or in the 2-p.p.b. date for each of the options considered below should be accurate.

Suggested halocarbon substitutes

The fundamental property of the substitute compounds that makes them more environmentally acceptable is a shorter atmospheric lifetime (<50 yr) so that they do not accumulate to large steady-state concentrations. Substitution levels assume a kg-per-kg replacement (in %) of the sum of halocarbon cuts referenced to their 1985 emission levels ($1,000 \text{ kt yr}^{-1}$ of CFCs, 80 kt yr^{-1} of CCl_4 , 140 kt yr^{-1} of HCFC-22, 600 kt yr^{-1} of CH_3CCl_3). The focus here is on stratospheric ozone rather than greenhouse warming and thus on substitutes containing chlorine. (Halon substitution by bromine compounds is not considered.) The International Technical Panel (part of the UNEP assessment¹²)

predicts 30% of the CFC market (350 kt yr^{-1} in 1985) could be replaced by short-lived substitutes containing chlorine, and that this market will grow at $3\% \text{ yr}^{-1}$ until terminated. Estimates from the chemical industry are similar¹⁸. Substitution for CCl_4 , HCFC-22 and CH_3CCl_3 —if these compounds were phased out—are likely not to involve chlorinated replacements. Hence, in the reference scenario with a complete cut in all halocarbons ($1,820 \text{ kt yr}^{-1}$ in 1985), the industry predictions would correspond to an initial substitution rate of about 20% in these scenarios.

Two surrogate compounds are chosen as arbitrary examples of possible replacements for the CFCs and are described in Table 1. Compound X has a lifetime of 15 yr (like HCFC-22), a single chlorine atom, a mean molecular weight of 115, and hence a steady-state chlorine loading factor of 0.10 relative to CFC-11. Compound Y has a lifetime of 6 yr (like CH_3CCl_3), a single chlorine atom, a mean molecular weight of 115, and hence a steady-state chlorine loading factor of 0.04 relative to CFC-11 (about one-third that of CH_3CCl_3). These substitutes are simply examples and do not correspond to any specific fluorocarbon¹⁴.

Environmental objectives

Several objectives must be considered when regulating the emissions of CFCs and halocarbons: (1) return chlorine loading to pre-industrial levels; (2) reduce chlorine abundances as soon as possible to those prevalent before the Antarctic ozone hole; and (3) minimize the peak chlorine loading of the atmosphere.

The pre-industrial atmosphere, insofar as chlorine is concerned, is not an achievable goal on a timescale of centuries, namely, the time to remove the present burden of CFCs from the atmosphere. Even in the reference scenario (complete phaseout of CFCs by the year 2000), chlorine levels remain above 1 p.p.b. until the year 2190. The more realistic goal (2) may be environmentally justified because the increase in atmospheric chlorine from pre-industrial levels up to the identification of the ozone hole was neither observed nor predicted to result

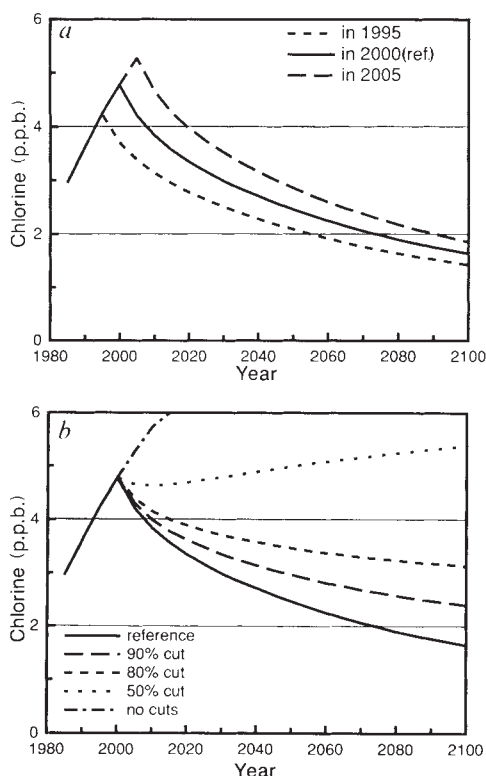


FIG. 1 a, Effect on chlorine concentrations of a 100% cut in all halocarbons; b, Partial cut in all halocarbons by year 2000.

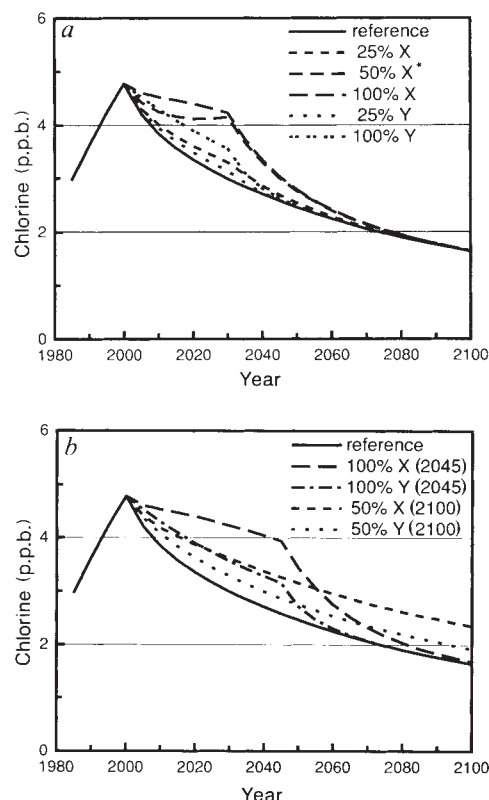


FIG. 2 Effect on chlorine concentrations from use of substitute compounds for (a) 30 years and (b) 45 years and beyond after the complete phaseout of halocarbons in the year 2000.

in decreases of column ozone greater than a few per cent³. Goal (3) may be viewed as reducing both global ozone depletion and the potential for an ozone hole in the Arctic. A number of scenarios are chosen to examine the possible options or controlling emissions that will affect the attainment of goals (2) and (3). **Option 1. Timing of the phaseout in CFCs, CCl₄, CH₃CCl₃ and HCFC-22.** We examine the environmental impact of eliminating the emissions of all halocarbons as a group (CFCs, CCl₄, CH₃CCl₃ and HCFC-22) by the end of years 1995, 2000 (reference case), and 2005, without substitution of any new chlorine-containing compounds. Figure 1a and Table 2 show that each five-year delay in the phaseout of these chemicals results in the peak chlorine loading increasing by about 0.5 p.p.b. (87% of it due to the CFCs) and in the date by which atmospheric chlorine will drop below 2 p.p.b. increasing by about 18 years. The critical factor determining the 2-p.p.b. date is the timing of the phaseout of the long-lived halocarbons (CFCs and CCl₄), not the short-lived halocarbons (CH₃CCl₃ and HCFC-22). Accumulation of the long-lived CFCs, as well as laxity in compliance with the phaseout (see options 2 and 4), make the date for effective cessation of CFC emissions slip further into the future.

Option 2 Incomplete phaseout of halocarbons. Here we examine the impact of a partial phaseout in the emissions of all current industrial halocarbons without any substitution of new chlorine-containing halocarbons. All reductions (90%, 80%, 50%) refer to 1985 emissions and occur at the end of year 2000; the scenario without reductions (0%) continues 1990 emissions until the year 2100.

Figure 1b shows that substantial although incomplete cutbacks in halocarbons (80–90%) can rapidly reduce chlorine loading. In these cases, atmospheric chlorine concentrations peak in the year in which the cuts are made, but the decrease in chlorine loading is slowed, and its asymptotic steady-state value is significantly raised by continued emissions at the 10–20% level. In the reference case (100% reduction), the 2-p.p.b. date is 2073 (see Table 2). With a 90% reduction, the chlorine loading reaches 2.4 p.p.b. in the year 2100 and falls below 2 p.p.b. after the year 2175. With an 80% reduction, the chlorine loading is 3.1 p.p.b. in the year 2100 and approaches a steady-state limit of ~2.8 p.p.b. With 50% reductions the chlorine loading drops briefly after the year 2000 and then rises above 5.3 p.p.b. by the year 2100. Without a cutback (0%) the chlorine abundance in the atmosphere rises rapidly, exceeding 8.5 p.p.b. by the year 2050 and 10.5 p.p.b. by the end of the twenty-first century.

Option 3 Substitution of alternative compounds X and Y. Here, the complete phaseout in the year 2000 (100% = 1,820 kt yr⁻¹) is followed by the use of substitute compounds X or Y at constant rates (25%, 100%) or at an initial rate (50%) that grows by a factor of 1.03 each year thereafter. The scenario with 3% yr⁻¹ growth (denoted by 50% X* in Fig. 2a) becomes equivalent to the constant 100% scenario by the year 2030 and would greatly exceed that level if continued. In general, we have limited the period of substitution to 30 years (2001–30), but in some cases it is extended to 45 years, or continued indefinitely. Results are shown in Figure 2. (The effect of a 45-yr substitution with Y is discussed below.)

The maximum chlorine loading occurs at the time of phaseout of CFCs for these cases. Modest substitution for the years 2001 to 2030 (25% with X or 50% with Y) still results in rapid decreases in chlorine loading after the year 2000; the added burden on top of the reference scenario is never more than 0.3 p.p.b.. In the case of 100% substitution with compound X, however, chlorine loading remains high, above 4 p.p.b., to the year 2032. Equivalent substitution with compound Y has a much smaller impact on atmospheric chlorine because of the shorter lifetime.

The added chlorine loading associated with substitution drops rapidly when emissions of these short-lived compounds cease; ~86% of the compound will be destroyed within two lifetimes (30 yr for X and 12 yr for Y). The 2-p.p.b. date is delayed by four years at most if compound X is phased out before the year

2030 or compound Y, before the year 2045. With indefinite substitution even at the 50% level, however, it is difficult, if not impossible, to achieve chlorine concentrations below 2 p.p.b. by the end of the twenty-first century.

Option 4 Delayed compliance of the phaseout by some producers. Here halocarbon emissions are cut substantially (80–90%), and the remaining fraction (20–10%) of their 1985 emissions are allowed to continue for 15 years until the end of year 2015. Substitution of compounds X and Y is considered as above, but with a 15-yr lag in fraction of emissions by the non-compliant producers (extending to the year 2045). Results are summarized in Figure 3 and Table 2.

An 80–90% cut (as in option 2) immediately reduces the peak chlorine loading in the year 2000. Furthermore, limiting the non-compliance to 15 yr results in only a 6 yr (90% cut) or 11 yr (80% cut) delay in the 2-p.p.b. date. When substitutes (X or Y) are added on top of the delayed phaseout, the largest increase in chlorine loading (relative to the corresponding substitution scenario in option 3) occurs just before the final halocarbon cuts (2015) and ranges from 0.2 p.p.b. (90% cut) to 0.4 p.p.b. (80% cut). In the case where 20% of the phaseout is delayed until the year 2015 and there is 100% replacement by compound X, the peak chlorine loading in the year 2000 (4.78 p.p.b.) is maintained out to the year 2015 (4.72 p.p.b.) and drops only slowly by the year 2030 (4.45 p.p.b.). Overall the greatest impact of this 15-yr delay in phaseout is to extend the duration of higher chlorine levels over the first quarter of the 21st century.

Option 5 A graduated phase-in of halocarbon cuts. Here we consider a two-step reduction in halocarbons (CFCs, CCl₄, CH₃CCl₃ and HCFC-22), first, to 50% of their 1985 levels at the end of year 1995 and, second, to a complete phaseout in the year 2000. This scenario is a hybrid of the 1995 and 2000 phaseouts in Option 1. Substitution is considered, also being

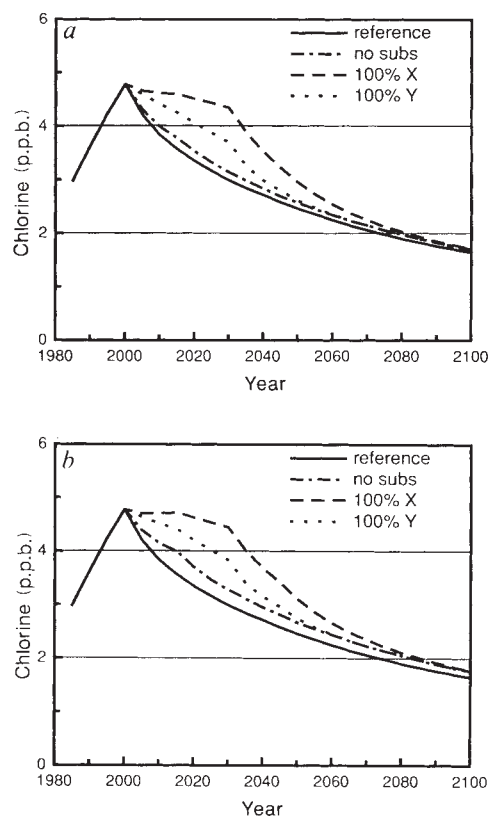


FIG. 3 Effect on chlorine concentrations of delaying the phaseout of final 10% of halocarbon emissions (a) and the final 20% of halocarbon emissions (b) by 15 yr.

phased in over 5 yr as with halocarbons. Results are given in Figure 4a and Table 2.

This complex scenario may be a more realistic outcome of current approaches to implement the Montreal Protocol and eliminate halocarbon emissions, recognizing that parts of the CFC market can phase out earlier, and including the effect of CFC accumulation in delaying release to the atmosphere. The history of chlorine abundance from this 50%/50% scenario lies between that of the year-1995 phaseout and that of the year-2000 phaseout. Peak chlorine loading occurs in 1995, and the 2-p.p.b. date is 2063, about 10 years earlier than in the reference case. Moderate substitutions (50% with X or Y) until the year 2030 give chlorine loadings in the latter half of the twenty-first century that are indistinguishable from the reference case.

Option 6 Continued use of CH_3CCl_3 . Methyl chloroform is a short-lived halocarbon (lifetime ~ 6 yr) with properties similar to some CFC alternatives; continued use would be equivalent to substitution with $1,800 \text{ kt yr}^{-1}$ of compound Y. Here we examine the impact of CH_3CCl_3 on chlorine loading by eliminating the emissions of CFCs, CCl_4 and HCFC-22 in the year 2000, but continuing CH_3CCl_3 emissions at 1990 levels until the year 2030. Results are summarized in Fig. 4b.

The phaseout of CH_3CCl_3 contributes to the rapid fall in chlorine loading after the year 2000 in the reference scenario, because the gas contributes about 0.5 p.p.b. to the total chlorine loading and is eliminated quickly from the atmosphere. However, continued use of CH_3CCl_3 for 30 yr gives no increase in the peak chlorine loading and little change in the 2-p.p.b. date. Owing to the accumulation of CFCs, a simultaneous cut in production of both CH_3CCl_3 and CFCs results in an almost immediate cessation of CH_3CCl_3 emissions and an extended period of CFC release. Under such circumstances, continued use of CH_3CCl_3 might lead to increases in the peak chlorine loading relative to the reference scenario, but this situation does not apply because the current CFC bank is not large enough.

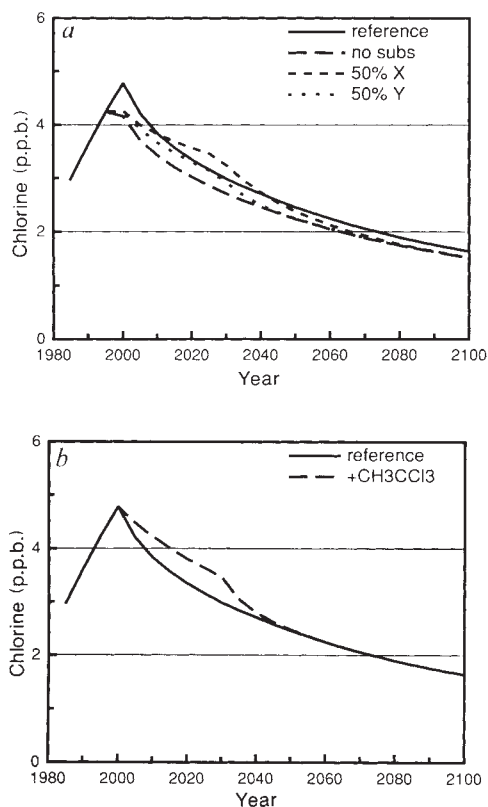


FIG. 4 a, Effect on chlorine concentrations of a two-step reduction in halocarbons: 50% cut in 1995 and complete phaseout at 2000; and b, effect of continuing methylchloroform emissions until 2030.

Option 7 Uncertainty in the atmospheric lifetimes of halocarbons.

Here we examine how predicted chlorine loading is affected by a $\pm 25\%$ uncertainty in the halocarbon lifetimes. Such ranges are extreme (for example, CFC-11 lifetimes vary from 45 to 75 yr) and cover the most probable range in CFC lifetimes from current atmospheric models and from budget analyses using observations and historical emissions¹⁹⁻²¹. Results are given in Fig. 5 and Table 2. When halocarbon lifetimes are reduced by 25%, the more rapid destruction of these compounds lowers the peak chlorine concentration by 0.3 p.p.b. (to 4.5 p.p.b.) and reduces the 2-p.p.b. date by 20 years (to the year 2053). When lifetimes increase by 25%, the peak chlorine concentration rises by 0.2 p.p.b. (to 5.0 p.p.b.) and extends the 2-p.p.b. date to the year 2095.

The uncertainty in predicting atmospheric chlorine concentrations following a halocarbon phaseout is not large, only a few tenths of a p.p.b. If these uncertainties in the atmospheric models are resolved in the near future, then the scenarios considered above (options 1-6) may need to be shifted as a group, but the calculated relative differences should be robust.

Option 8 The impact of halon emissions on bromine loading. The change in bromine loading is calculated from the emissions of the halons 1301 (CF_3Br) and 1211 (CF_2ClBr); halon 2402 ($\text{CF}_2\text{BrCF}_2\text{Br}$) is assumed to be negligible at present and in the future. (Short-lived compounds such as bromoform, CHBr_3 , are expected to contribute little to stratospheric bromine.) This scenario is similar to that for chlorine loading and examines the effects of cutting halon emissions by 100%, 80% and 50% of their 1985 levels, as well as freezing emissions (0% cut) at their 1990 levels. Results are shown in Fig. 6.

In 1985 the halons 1301 and 1211 together contributed about one-sixth of the total bromine loading. If their emissions

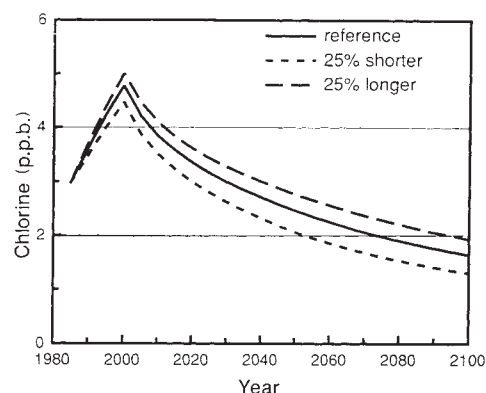


FIG. 5 Sensitivity of chlorine concentrations to uncertainties in the atmospheric lifetimes of CFCs and other halocarbons.

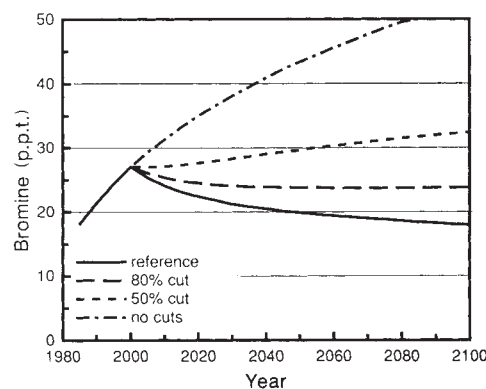


FIG. 6 Effect on bromine concentrations of cuts in halon emissions.

continued at the estimated 1990 production levels then the atmospheric abundance of bromine in the future would approximately triple from about 18 p.p.t. (parts per 10^{12}) (3 p.p.t. halons) in the year 1985 to 53 p.p.t. (38 p.p.t. halons) by the year 2100. In these calculations CH_3Br remains constant; but observations²² suggest that large sources are associated with northern continents and human activity (for example grain fumigation).

To evaluate the impact of bromine increases, it would be useful to define a factor relating bromine-catalysed destruction of ozone to that of chlorine. Such a simple factor is difficult to derive because ozone depletion is occurring over a wide range of photochemical environments, ranging from global scales in the upper stratosphere (near 40 km altitude where bromine contributes little to the restricted environment of the Antarctic ozone hole (near 20 km altitude where BrO is part of a key catalytic cycle). Further, much of the bromine-catalysed loss depends on the ClO abundance. One can estimate this factor to be about 30 (refs 7–10) with a large uncertainty ranging from 10 to 50. With a freeze on halon emissions (0% cut) the increase in bromine loading by the year 2100 is ~ 35 p.p.t., or the equivalent of adding ~ 1 p.p.b. of chlorine to the atmosphere.

Conclusions

We must constrain our emissions of chlorine- and bromine-containing compounds to reverse the ozone depletion that has occurred to date and to reduce the potential for even greater ozone losses in the future. These objectives require a truly global effort and cannot be achieved by the action of a single industry or group of nations. A phaseout of almost all emissions of halocarbons is needed in the next century in order to decrease

the atmospheric abundance of chlorine below 2 p.p.b. before the year 2100 (the minimum necessary for recovery of the Antarctic ozone hole). Each year's delay means an extra 0.1 p.p.b. in the peak chlorine loading and pushes the 2-p.p.b. date 3.6 years further into the future.

Ideally, we should cease emissions of CFCs and other halocarbons immediately. All other options result in enhanced levels of stratospheric chlorine and bromine sometime in the future. In reality, we may need to accept compromises to achieve global cooperation with a halocarbon phaseout early in the next century. We have therefore examined possible options of delayed and partial compliance as well as substitution of long-lived CFCs with short-lived HCFCs. The purpose of these scenarios is to identify options that do not substantially enhance the risk of greater ozone depletion.

The environmental goals outlined above might best be achieved if we phase out production/emission of long-lived halocarbons (CFCs and CCl_4) as soon as possible; achieve as great as possible compliance with an immediate phaseout of CFCs and CCl_4 at the expense of substituting short-lived HCFCs; allow a 15-year lag in implementing the CFC phaseout by a small fraction (10–20%) of the community if it prevents long-term non-compliance; make significant cutbacks in the emissions of abundant, short-lived halocarbons, particularly CH_3CCl_3 , if it becomes necessary to reduce rapidly the peak chlorine loading; phase out the chlorine-containing halocarbon substitutes sometime by the middle of the next century; and stabilize stratospheric bromine levels by greatly reducing (by 80% or more) the production of halons, especially the long-lived CF_3Br . □

Received 17 October 1989; accepted 5 February 1990.

1. Montreal Protocol on Substances that Deplete the Ozone Layer, Final Act, UNEP (1987).
2. Molina, L. T. & Molina, M. J. *J. phys. Chem.* **91**, 433–436 (1986).
3. Wofsy, S. C., McElroy, M. B. & Yung, Y. L. *Geophys. Res. Lett.* **2**, 215–218 (1975).
4. Prather, M. J., McElroy, M. B. & Wofsy, S. C. *Nature* **312**, 227–231 (1984).
5. McElroy, M. B., Salawitch, R. J., Wofsy, S. C. & Logan, J. A. *Nature* **321**, 759–762 (1986).
6. Anderson, J. G. *et al. J. geophys. Res.* **94**, 11480–11520 (1989).
7. Ko, M. K. W. *et al. J. geophys. Res.* **94**, 16705–16716 (1989).
8. Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207–210 (1985).
9. Stolarski, R. S. *et al. Nature* **322**, 808–811 (1986).
10. Krueger, A. J., Schoeberl, M. R., Stolarski, R. S. & Sechrist, F. S. *Geophys. Res. Lett.* **15**, 1365–1368 (1988).
11. Watson, R. T., Prather, M. J. & Kurylo, M. J. *NASA Reference Publication* **1208**, 1–208 (1988).
12. *Scientific Assessment of Stratospheric Ozone: 1989 World Meteorological Organization, Global Ozone Research and Monitoring Project Rep. No. 20* (WMO, Geneva, 1990).
13. Fisher, D. A. *et al. Nature* **344**, 508–512 (1990).

14. AFEAS, *Alternative Fluorocarbon Environmental Acceptability Study, Scientific Assessment of Stratospheric Ozone: 1989, Vol. II Appendix: AFEAS*, World Meteorological Organization, Global Ozone Research and Monitoring Project, Rep. No. 20 (WMO, Geneva, 1990).
15. Heidt, L. E. *et al. J. geophys. Res.* **94**, 11599–11612 (1989).
16. Podolske, J. R., Lowenstein, M., Strahan, S. E. & Chan, K. R. *J. geophys. Res.* **94**, 17767–17772 (1989).
17. *1987 Production and Sales of Chlorofluorocarbons CFC-11 and CFC-12* (Chemical Manufacturers Association, Fluorocarbon Program Panel, Washington DC, 1988).
18. *DuPont Magazine* Vol 83, No. 5 (1989).
19. Cunnold, D. M. *et al. J. geophys. Res.* **91**, 10797–10817 (1986).
20. Prinn, R. G. *et al. Science* **239**, 945–950 (1987).
21. Rowland, F. S. *et al. Geophys. Res. Lett.* **9**, 481–484 (1982).
22. Perket, S. A., Jones, B. M. R., Rycroft, M. J. & Simmons, D. A. *Nature* **318**, 550–553 (1985).

ACKNOWLEDGEMENTS. We emphasize that this analysis is based on the substantial record of research from the atmospheric sciences community, and, in particular, it relies on those who have contributed over the years to the international ozone assessments.

Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series

George Sugihara* & Robert M. May†

* Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA

† Department of Zoology, Oxford University, Oxford, OX1 3PS, UK

An approach is presented for making short-term predictions about the trajectories of chaotic dynamical systems. The method is applied to data on measles, chickenpox, and marine phytoplankton populations, to show how apparent noise associated with deterministic chaos can be distinguished from sampling error and other sources of externally induced environmental noise.

TWO sources of uncertainty in forecasting the motion of natural dynamical systems, such as the annual densities of plant or animal populations, are the errors and fluctuations associated with making measurements (for example, sampling errors in estimating sizes, or fluctuations associated with unpredictable environmental changes from year to year), and the complexity of the dynamics themselves (where deterministic dynamics can easily lead to chaotic trajectories).

Here we combine some new ideas with previously developed techniques^{1–7,16,24–26}, to make short-term predictions that are

based on a library of past patterns in a time series¹. By comparing the predicted and actual trajectories, we can make tentative distinctions between dynamical chaos and measurement error: for a chaotic time series the accuracy of the nonlinear forecast falls off with increasing prediction-time interval (at a rate which gives an estimate of the Lyapunov exponent²), whereas for uncorrelated noise, the forecasting accuracy is roughly independent of prediction interval. For a relatively short time series, distinguishing between autocorrelated noise and chaos is more difficult; we suggest a way of distinguishing such 'coloured' noise from chaos in our scheme, but questions remain, at least for time series of finite length.

The method also provides an estimate of the number of dimensions, or 'active variables', of the attractor underlying a time series that is identified as chaotic. Unlike many current approaches to this problem (for example, that of Grassberger and Procaccia⁸), our method does not require a large number of data points, but seems to be useful when the observed time series has relatively few points (as is the case in essentially all ecological and epidemiological data sets).

Forecasting for a chaotic time series

Below, we outline the method and show how it works by apply-

ing it to a chaotic time series generated artificially from the deterministic 'tent map'. We then apply it to actual data on measles and chickenpox in human populations (which have been previously analysed using different techniques⁹⁻¹³) and on diatom populations. We conclude that the method may be capable of distinguishing chaos from measurement error even in such relatively short runs of real data.

As an example of the difficulties in short-range forecasting, we consider the chaotic time series shown in Fig. 1a. This time series was generated from the first-difference transformation ($x_{t+1} - x_t$) on the deterministic tent map or triangular 'return map' (described in detail in the legend to Fig. 1a). Here and elsewhere in this report, we first-difference the data partly to give greater density in phase space to such chaotic attractors as may exist, and partly to clarify nonlinearities by reducing the effects of any short-term linear autocorrelations. It should be noted, however, that both in our artificial examples and in our later analysis of real data, we obtain essentially the same results if we work with the raw time series (without first-differencing). With the exception of a slight negative correlation between immediately adjacent values, the sequence in Fig. 1a is uncorrelated, and is in many ways indistinguishable from white noise: the null hypothesis of a flat Fourier spectrum cannot be rejected

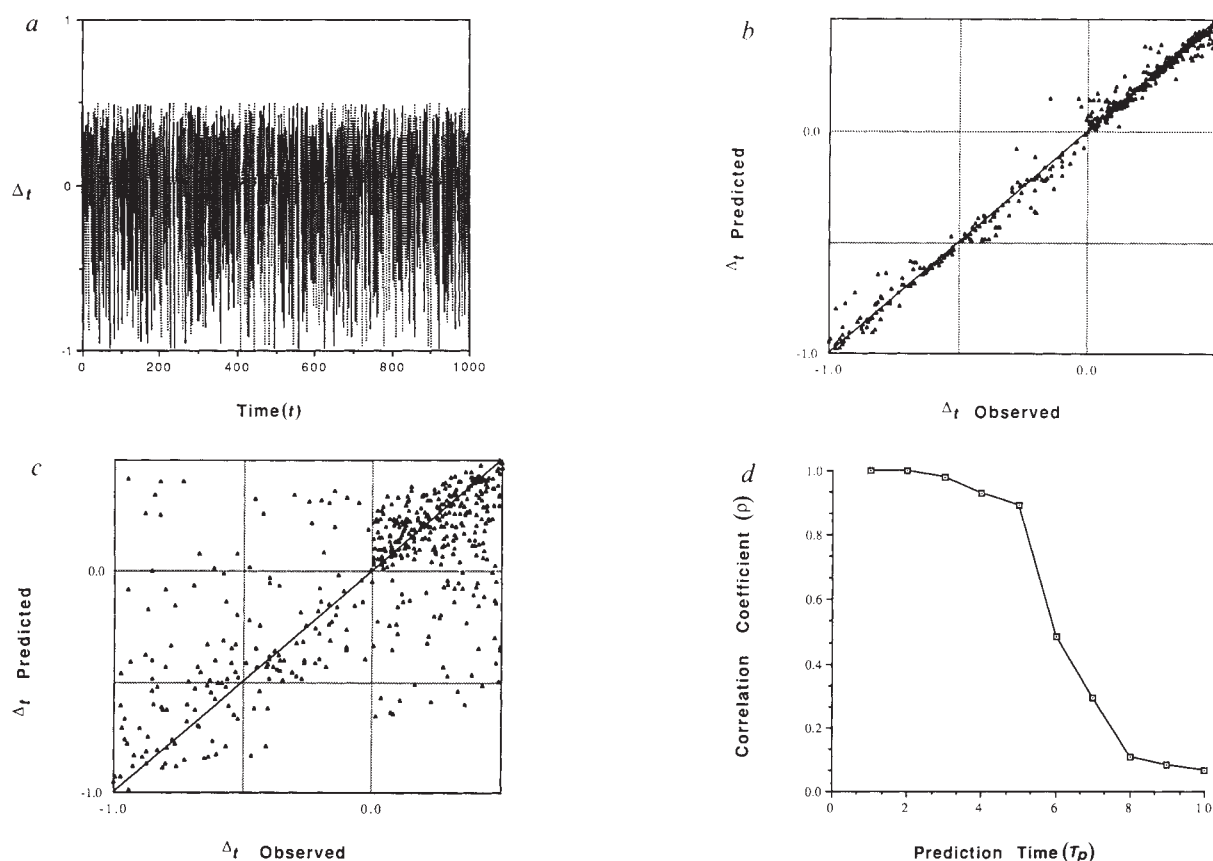


FIG. 1 a, Time series of 1,000 points (which in many ways is indistinguishable from white noise) generated by taking first-differences, $\Delta_t = x_{t+1} - x_t$, of the tent map: $x_{t+1} = 2x_t$ for $0.5 > x_t > 0$; $x_{t+1} = 2 - 2x_t$ for $1 > x_t > 0.5$. b, Predicted values two steps into the future ($T_p = 2$) versus observed values for the tent delta time series depicted in a. Specifically, the first 500 points in the series were used to generate a library of patterns, which were then used as a basis for making predictions for each of the second 500 points. As described in the text, the predictions were made using a simplex projection method, and in this figure the embedding dimension and lag time are $E = 3$ and $\tau = 1$, respectively. Here the coefficient of correlation between predicted and actual values is $\rho = 0.997$ ($N = 500$). For comparison, we note that the corresponding correlation coefficient obtained using the first half of the series to predict the second half with an autoregressive linear model

(where the predictions are based on the weighted average of three linear maps, one for each of the three different τ -values that give the best results in such a linear scheme) is $\rho = 0.04$. c, Exactly as for Fig. b, except here the predictions are five time steps into the future ($T_p = 5$). The correlation coefficient between predicted and actual values is now $\rho = 0.89$ ($N = 500$). d, Summary of the trend between b and c, by showing ρ between predicted and observed values in the second half (second 500 points) of the time series of a, as a function of T_p . As in b and c, the simplex projection method here uses $E = 3$ and $\tau = 1$. That prediction accuracy (as measured by the coefficient of correlation between predicted and observed values) falls as predictions extend further into the future is a characteristic signature of a chaotic attractor.

using Bartlett's Kolmogorov-Smirnov test, with $P=0.85$. Because nonadjacent values in the time series are completely uncorrelated, standard statistical methods (that is, linear autoregression) cannot be used to generate predictions two or more steps into the future that are significantly better than the mean value (that is, zero) for the series.

Figure 1b and c show the results of local forecasting with the above data. The basic idea here, as outlined below, is that if deterministic laws govern the system, then, even if the dynamical behaviour is chaotic, the future may to some extent be predicted from the behaviour of past values that are similar to those of the present.

Specifically, we first choose an 'embedding dimension', E , and then use lagged coordinates to represent each lagged sequence of data points $\{x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(E-1)\tau}\}$ as a point in this E -dimensional space; for this example we choose $\tau=1$, but the results do not seem to be very sensitive to the value of τ , provided that it is not large^{14,15}. For our original time series, shown in Fig. 1a, each sequence for which we wish to make a prediction—each 'predictee'—is now to be regarded as an E -dimensional point, comprising the present value and the $E-1$ previous values each separated by one lag time τ . We now locate all nearby E -dimensional points in the state space, and choose a minimal neighbourhood defined to be such that the predictee is contained within the smallest simplex (the simplex with minimum diameter) formed from its $E+1$ closest neighbours; a simplex containing $E+1$ vertices (neighbours) is the smallest simplex that can contain an E -dimensional point as an interior point (for points on the boundary, we use a lower-dimensional simplex of nearest neighbours). The prediction is now obtained by projecting the domain of the simplex into its range, that is by keeping track of where the points in the simplex end up after p time steps. To obtain the predicted value, we compute where the original predictee has moved within the range of this simplex, giving exponential weight to its original distances from the relevant neighbours. This is a nonparametric method, which uses no prior information about the model used to generate the time series, only the information in the output itself. It should apply to any stationary or quasi-ergodic dynamic process, including chaos. This method is a simpler variant of several more complicated techniques explored recently by Farmer and Sidorowich³ and by Casdagli⁷.

Figure 1b compares predicted with actual results, two time steps into the future. Figure 1c makes the same comparison, but at five time steps into the future. There is obviously more scatter in Fig. 1c than in Fig. 1b. Figure 1d quantifies how error increases as we predict further into the future in this example, by plotting the conventional statistical coefficient of correlation, ρ , between predicted and observed values as a function of the prediction-time interval, T_p (or the number of time steps into the future, p). Such decrease in the correlation coefficient with increasing prediction time is a characteristic feature of chaos (or equivalently, of the presence of a positive Lyapunov exponent, with the magnitude of the exponent related to the rate of decrease of ρ with T_p). This property is noteworthy, because it indicates a simple way to differentiate additive noise from deterministic chaos: predictions with additive noise that is uncorrelated (in the first-differences) will seem to have a fixed amount of error, regardless of how far, or close, into the future one tries to project, whereas predictions with deterministic chaos will tend to deteriorate as one tries to forecast further into the future. Farmer and Sidorowich^{3,16} have derived asymptotic results (for very long time series, $N \gg 1$) that describe how this error typically propagates, over time, in simple chaotic systems. The standard correlation coefficient is one of several alternative measures of the agreement between predicted and observed values; results essentially identical to those recorded in Figs 1–6 can be obtained with other measures (such as the mean squared difference between predicted and observed values as a ratio to the mean squared error).

Forecasting with uncorrelated noise

Figure 2a (solid line) shows that, indeed, this signature of ρ decreasing with T_p does not arise when the erratic time series is in fact a noisy limit cycle. Here we have uncorrelated additive noise superimposed on a sine wave (Fig. 2b). Such uncorrelated noise is reckoned to be characteristic of sampling variation.

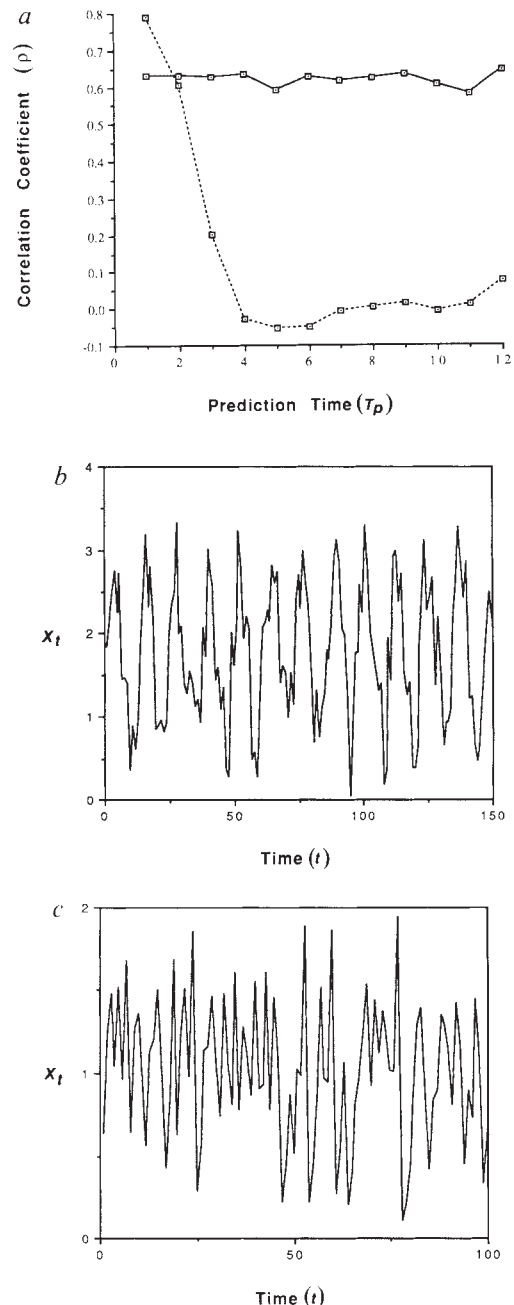


FIG. 2 a, Solid line shows ρ between predicted and observed values for the second half of the time series defined in b (which is, in fact, a sine wave with additive noise) as a function of T_p . As discussed in the text, the accuracy of the prediction, as measured by ρ , shows no systematic dependence on T_p . By contrast, the time series shown in c (which is the sum of two separate tent map series) does show the decrease in ρ with increasing T_p , as illustrated by the dashed line, that is characteristic of a chaotic sequence. Both curves are based on the simplex methods described in the text, with $E=3$ and $\tau=1$. b, First 150 points in the time series generated by taking discrete points on a sine wave with unit amplitude ($x_t = \sin(0.5t)$), and adding a random variable chosen (independently at each step) uniformly from the interval $[-0.5, 0.5]$. That is, the series is generated as a 'sine wave + 50% noise'. c, Time series illustrated here is generated by adding together two independent tent map sequences.

Here the error remains constant as the simplex is projected further into the future; past sequences of roulette-wheel numbers that are similar to present ones tell as much or little about the next spin as the next hundredth spin. By contrast, the dashed line in Fig. 2a represents ρ as a function of T_p , for a chaotic sequence generated as the sum of two independent runs of tent map; that is, for the time series illustrated in Fig. 2c. Although the two time series in Fig. 2b and Fig. 2c can both look like the sample functions of some random process, the characteristic signatures in Fig. 2a distinguish the deterministic chaos in Fig. 2c from the additive noise in Fig. 2b.

The embedding dimension

The predictions in Figs 1 and 2 are based on an embedding dimension of $E=3$. The results are, however, sensitive to the choice of E . Figure 3a compares predicted and actual results for the tent map two time steps ahead ($T_p=2$), as in Fig. 1b, except that now $E=10$ (versus $E=3$ in Fig. 1b). Clearly the predictions are less accurate with this higher embedding dimension. More generally, Fig. 3b shows ρ between predicted and actual results one time step into the future ($T_p=1$) as a function of E , for two different choices of the lag time ($\tau=1$ and $\tau=2$). It may seem surprising that having potentially more information—more data summarized in each E -dimensional point, and a higher-dimensional simplex of neighbours of the predictee—reduces the accuracy of the predictions; in this respect, these results differ from results reported by Farmer and Sidorowich for parametric forecasting involving linear interpolation to construct local polynomial maps³. We think this effect is caused by contamination of nearby points in the higher-dimensional embeddings with points whose earlier coordinates are close, but whose recent (and more relevant) coordinates are distant. If this is so, our method may have additional applications as a trial-and-error method of computing an upper-bound

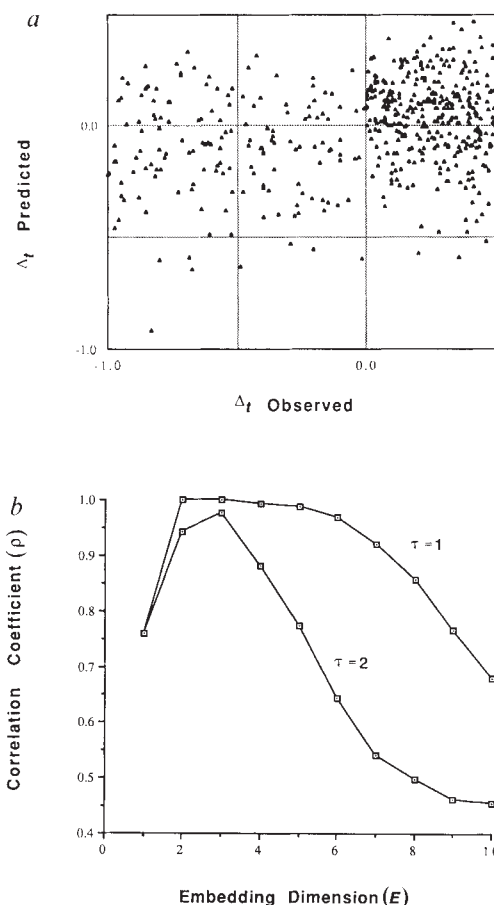
on the embedding dimension, and thence on the dimensionality of the attractor (see also refs 2, 6, 7).

Problems and other approaches

We have applied these ideas to a variety of other 'toy models', including the quadratic map along with other first-order difference equations and time series obtained by taking points at discrete time intervals from continuous chaotic systems such as those of the Lorenz and Rossler models (in which the chaotic orbits are generated by three coupled, nonlinear differential equations). The results for ρ as a function of T_p are in all cases very similar to those shown in Fig. 1d. Even in more complicated cases, such as those involving the superposition of different chaotic maps, we observe a decline in ρ versus T_p ; here, however, the signature can show a step pattern, with each step corresponding to the dominant Lyapunov exponent for each map.

So far, we have compared relationships between ρ and T_p for chaotic time series with the corresponding relations for white noise. More problematic, however, is the comparison with ρ - T_p relationships generated by coloured noise spectra, in which there are significant short-term autocorrelations, although not long-term ones. Such autocorrelated noise can clearly lead to correlations, ρ , between predicted and observed values that decrease as T_p lengthens. Indeed, it seems likely that a specific pattern of autocorrelations could be hand-tailored, to mimic any given relationship between ρ and T_p (such as that shown in Fig. 1d) obtained from a finite time series. We conjecture, however, that such an artificially designed pattern of autocorrelation would in general give a flatter ρ -versus- E relationship than those of simple chaotic time series corresponding to low-dimensional attractors (for example, see Fig. 3b). Working from the scaling relations for error versus T_p in chaotic systems^{3,16}, Farmer (personal communication) has indeed suggested that asymptotically (for very large N), the ρ - T_p relationships generated by

FIG. 3 a, Similar to Fig. 1b, this figure shows predictions one time step into the future ($T_p=1$) versus observed values, for the second 500 points in the tent map time series of Fig 1a, with the difference that here we used an embedding dimension $E=10$ (in contrast to $E=3$ in Fig. 1b; the lag time remains unchanged at $\tau=1$). As discussed in the text, the accuracy of the prediction deteriorates as E gets too large ($\rho=0.25$, $N=500$). b, Correlation coefficient between predicted and observed results, ρ , is shown as a function of E for predictions one time step into the future ($T_p=1$). The relationship is shown for $\tau=1$ and $\tau=2$. The figure indicates how such empirical studies of the relation between ρ and E may be used to assess the optimal E .



autocorrelated noise may characteristically scale differently from those generated by deterministic chaos. Although we have no solution to this central problem—which ultimately may not have any general solution, at least for time series of the sizes found in population biology—we suggest that an observed time series may tentatively be regarded as deterministically chaotic if, in addition to a decaying ρ - T_p signature, the correlation, ρ , between predicted and observed values obtained by our methods is significantly better than the corresponding correlation coefficient obtained by the best-fitting autoregressive linear predictor (see also, ref. 16). For the tent map, as detailed in the legend to Fig. 1, *b* and *c*, our nonlinear method gives ρ values significantly better than those from autoregressive linear models (composed of the weighted average of the three best linear maps).

Most previous work applying nonlinear theory to experimental data begins with some estimate of the dimension of the underlying attractor²⁻⁷. The usual procedure (for exceptions, see refs 2, 6, 7, 25) is to construct a state-space embedding for

the time series, and then to calculate the dimension of the putative attractor using some variant of the Grassberger-Procaccia algorithm⁸. A correlation integral is calculated that is essentially the number of points in E space separated by a distance less than l , and the power-law behaviour of this correlation integral (l^ν) is then used to estimate the dimension, D , of the attractor ($D \geq \nu$). This dimension is presumed to give a measure of the effective number of degrees of freedom or 'active modes' of the system. An upper bound on a minimal embedding dimension (which can be exceeded when the axes of the embedding are not truly orthogonal) is $E_{\min} < 2D + 1$, where D is the attractor dimension^{8,15}. The scaling regions used to estimate power laws by these methods are typically small and, as a consequence, such calculations of dimension involve only a small fraction of the points in the series (that is, they involve only a small subset of pairs of points in the state space). In other words, the standard methods discard much of the information in a time series, which, because many natural time series are of limited size, can be a serious problem. Furthermore, the

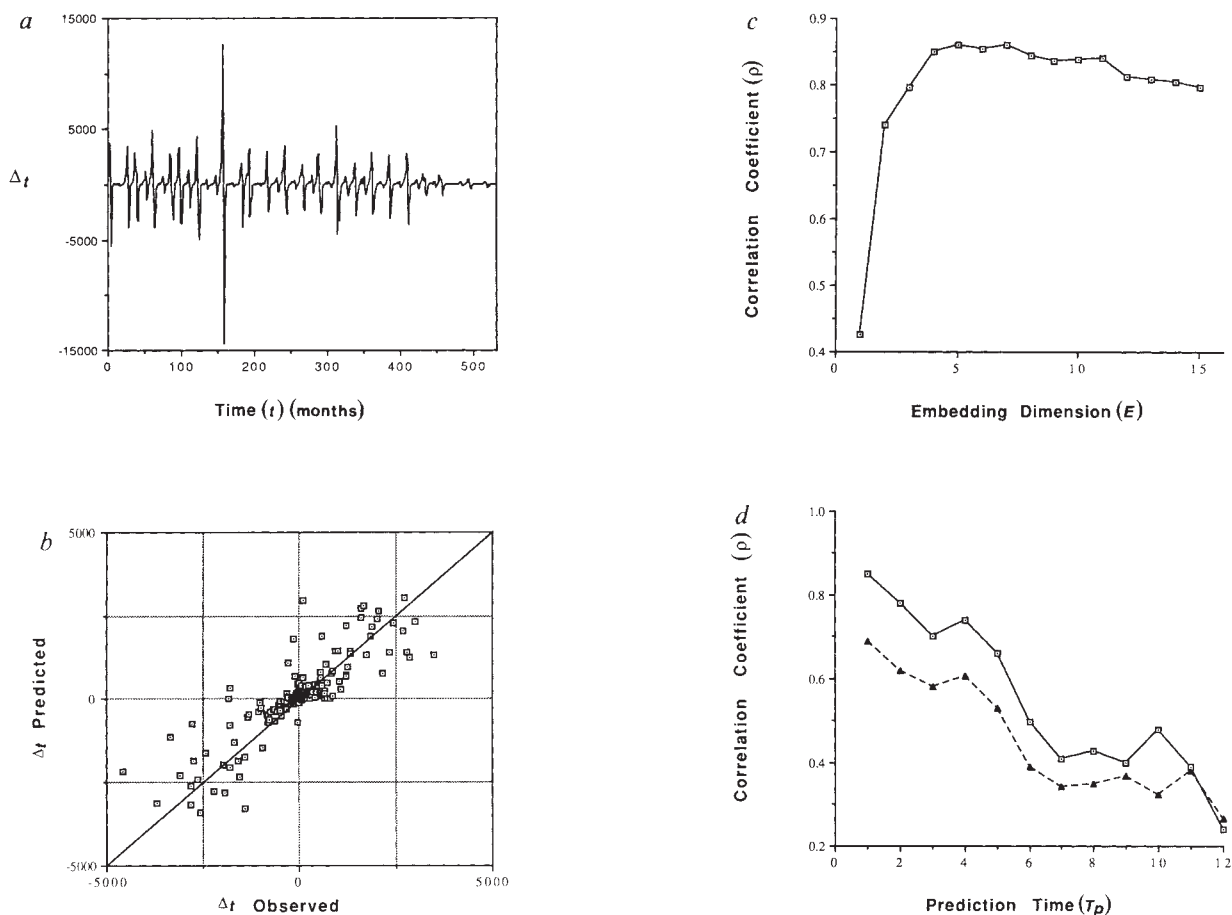


FIG. 4 *a*, Time series generated by taking first differences, $x_{t+1} - x_t$, of the monthly number of cases of measles reported in New York City between 1928 and 1972 (the first 532 points in the sequence shown here). After 1963, the introduction of immunization against measles had a qualitative effect on the dynamics of infection; this can be seen in the later part of the sequence illustrated here. *b*, Using the methods described earlier, the first part of the series in Fig. 4*a* (216 points, 1928 to 1946) was used to construct a library of past patterns, which were then used as a basis for predicting forward from each point in the second part of the series, from 1946 to 1963. Predicted and observed values are shown here for predictions one time step into the future ($T_p=1$), using $E=6$ and $\tau=1$. The correlation coefficient between predicted and observed values is $\rho=0.85$ ($P<10^{-5}$ for $N=216$). For comparison, the corresponding prediction based on an autoregressive linear model (composed of five optimal linear maps, compare Fig. 1*b*) gives $\rho=0.72$ (which is significantly different from $\rho=0.85$ at the

$P<0.0005$ level). *c*, As in Fig. 3*b*, ρ between predicted and observed results, is shown as a function of E (for $T_p=1$ and $\tau=1$). This figure indicates an optimal embedding dimension of $E \sim 5-7$, corresponding to a chaotic attractor with dimension 2-3. *d*, Here ρ , between predicted and observed results for measles, is shown as a function of T_p (for $E=6$ and $\tau=1$). For the points connected by the solid lines, the predictions are for the second half of the time series (based on a library of patterns compiled from the first half). For the points connected by the dashed lines, the forecasts and the library of patterns span the same time period (the first half of the data). The similarity between solid and dashed curves indicates that secular trends in underlying parameters do not introduce significant complications here. The overall decline in prediction accuracy with increasing time into the future is, as discussed in the text, a signature of chaotic dynamics as distinct from uncorrelated additive noise.

Grassberger-Procaccia and related methods are somewhat more qualitative, requiring subjective judgement about whether there is an attractor of given dimensions. Prediction methods, by contrast, have the advantage that standard statistical criteria can be used to evaluate the significance of the correlation between predicted and observed values. As Farmer and Sidorowich^{3,16}, and Casdagli⁷, have also emphasized, prediction methods should provide a more stringent test of underlying determinism in situations of given complexity. Prediction is, after all, the *sine qua non* of determinism.

Time series from the natural world

Measles. For reported cases of measles in New York City, there is a monthly time series extending from 1928 (ref. 17). After 1963, immunization began to alter the intrinsic dynamics of this system, and so we use only the data from 1928 to 1963 ($N = 432$). These particular data have received a lot of attention recently, and they are the focus of a controversy about whether the dynamics reflect a noisy limit cycle⁹ or low-dimensional chaos superimposed on a seasonal cycle¹⁰⁻¹³. In particular, the data have been carefully studied by Schaffer and others^{13-16,27}, who have tested for low-dimensional chaos using a variety of methods, including the Grassberger-Procaccia algorithm¹³, estimation of Lyapunov exponents¹³, reconstruction of Poincaré return maps^{10,11}, and model simulations^{12,13}. Although it is not claimed that any of these tests are individually conclusive,

together they support the hypothesis that the measles data are described by a two- to three-dimensional chaotic attractor.

Figure 4a shows the time series obtained by taking first differences, $X_{t+1} - X_t$, of these data. As discussed above, the first difference was taken to 'whiten' the series (that is, reduce autocorrelation) and to diminish any signals associated with simple cycles (a possibility raised by proponents of the additive noise hypothesis⁹). We then generated our predictions by using the first half of the series (216 points) to construct an ensemble of points in an E -dimensional state space, that is, to construct a library of past patterns. The resulting information was then used to predict the remaining 216 values in the series, along the lines described above, for each chosen value of E . Figure 4b, for example, compares predicted and observed results, one time step into the future ($T_p = 1$ month), with $E = 6$. Figure 4c shows ρ between predicted and observed results as a function of E for $T_p = 1$. Taking the optimal embedding dimension to be that yielding the highest correlation coefficient (or least error) between prediction and observation in one time step, it is seen from Fig. 4c that $E \approx 5-7$. This accords with previous estimates¹⁰⁻¹³ made using various other methods, and is consistent with the finding of an attractor with dimension $D = 2-3$.

The points joined by the solid lines in Fig. 4d show ρ as a function of T_p (for $E = 6$). Prediction error seems to propagate in a manner consistent with chaotic dynamics. This result, in combination with the significantly better performance ($P <$

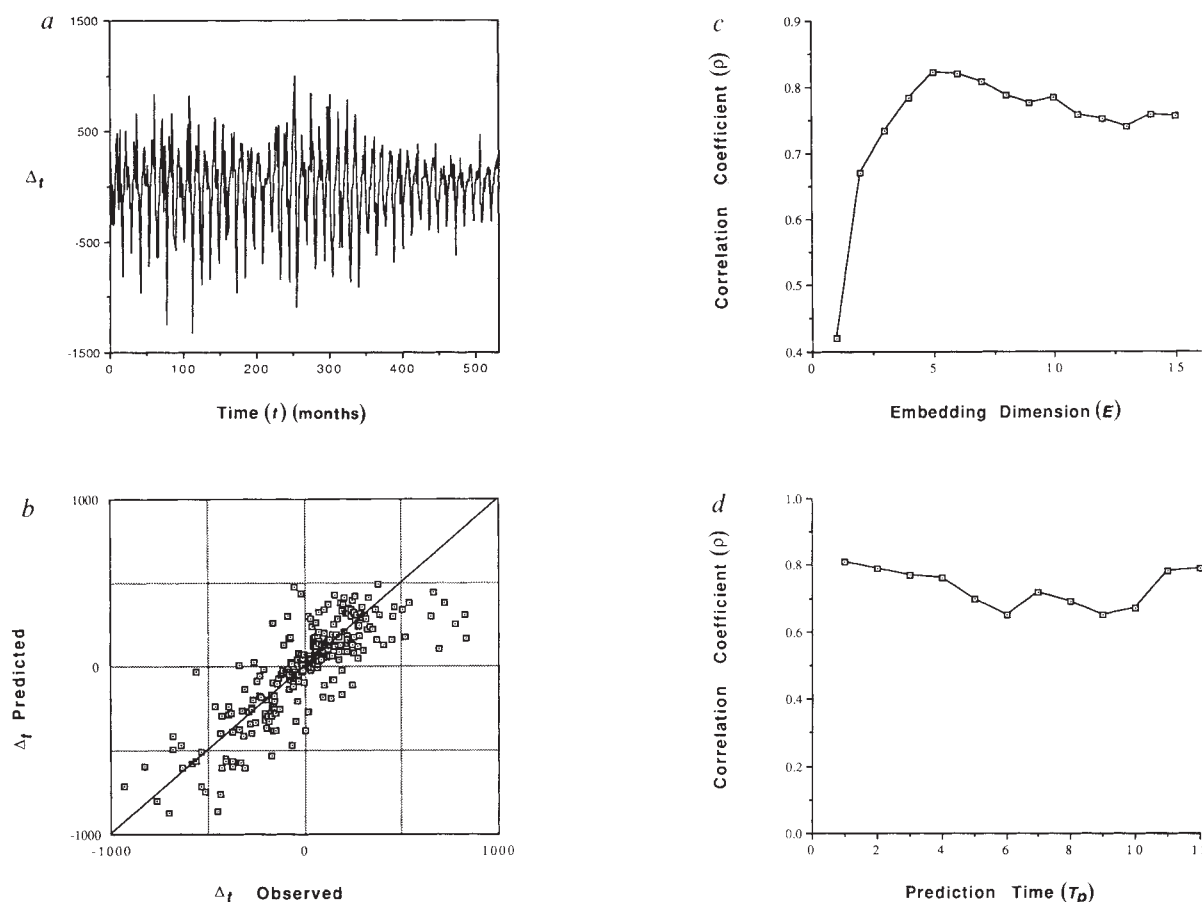


FIG. 5 a, As for Fig. 4a, except the time series comes from taking first-differences of the monthly numbers of reported cases of chickenpox in New York City from 1928 to 1972. b, As in Fig. 4b, predicted and observed numbers of cases of chickenpox are compared, the predictions being one time step into the future, $T_p = 1$ (here, $E = 5$ and $\tau = 1$). The correlation coefficient between predicted and observed values is $\rho = 0.82$; an autoregressive linear model alternatively gives predictions which have $\rho = 0.84$. In contrast to Fig. 4b for measles, here there is no significant difference

between our prediction technique and standard linear autoregressive methods. c, Correlation coefficient between predicted and observed results for chickenpox, ρ , shown as a function of E for predictions one time step into the future ($T_p = 1$ and $\tau = 1$). d, Compare with Fig. 4d; ρ , between predicted and observed values, as a function of T_p (with $E = 5$ and $\tau = 1$) is shown. Here the lack of dependence of ρ on T_p , which is in marked contrast with the pattern for measles in Fig. 4d, indicates pure additive noise (superimposed on a basic seasonal cycle).

0.0005) of our nonlinear predictor as compared with an optimal linear autoregressive model (see legend to Fig. 4b) agrees with the conclusion that the noisy dynamics shown in Fig. 4a are, in fact, deterministic chaos¹⁰⁻¹³.

For data from the natural world, as distinct from artificial models, physical or biological parameters, or both, can undergo systematic changes over time. In this event, libraries of past patterns can be of dubious relevance to an altered present and even-more-different future. In a different context, there is the example of how secular trends in environmental variables can complicate an analysis of patterns of fluctuation in the abundance of bird species^{18,19}. We can gauge the extent to which secular trends might confound our forecasting methods in the following way. Rather than using the first half of the time series to compile the library of patterns, and the second half to compute correlations between predictions and observations, we instead investigate the case in which the library and forecasts span the same time period. Therefore we focus our predictions in the first half of the series, from which the library was drawn. To avoid redundancy, however, between our forecasts and the model, we sequentially exclude points from the library that are in the neighbourhood of each predictee (specifically, the $E\tau$ points preceding and following each forecast). The points con-

nected by the dashed lines in Fig. 4d show the ρ versus T_p relationship that results from treating the measles data in this way (again with $E=6$). The fairly close agreement between these results (for which the library of patterns and the forecasts span the same time period) and those of the simpler previous analysis (the solid line in Fig. 4d) indicates that within these time frames, secular trends in underlying parameters are not qualitatively important.

Chickenpox. Figure 5a-d repeat the process just described for measles, but now for monthly records of cases of chickenpox in New York City from 1949 to 1972 (ref. 20). Figure 5a shows the time series of differences, $X_{t+1} - X_t$. The 532 points in Fig. 5a are divided into two halves, with the first half used to construct the library, on which predictions are made for the second 266 points. These predictions are compared with the actual data points, as shown for predictions one time step ahead ($T_p=1$ month) in Fig. 5b. In Fig. 5b, $E=5$; Fig. 5c shows that an optimum value of E , in the sense just defined, is about 5 to 6. By contrast with Fig. 4d for measles, ρ between predicted and observed results for chickenpox shows no dependence on T_p : one does as well at predicting the incidence next year as next month. Moreover, the optimal linear autoregressive model

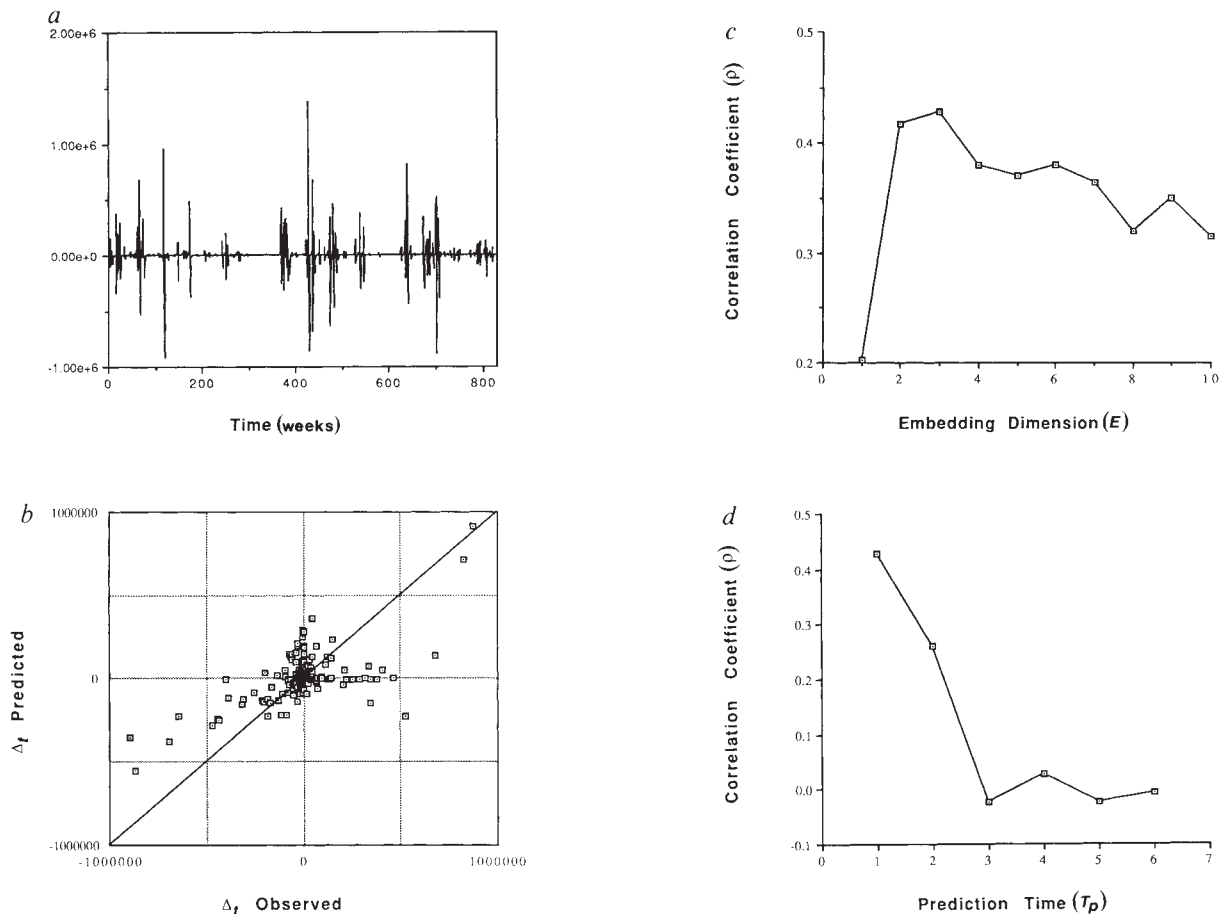


FIG. 6 a, Time series of first differences, $x_{t+1} - x_t$, of the weekly numbers of diatoms in seawater samples taken at Scripps Pier, San Diego, from 1929 to 1939 ($N=830$). b, Using the first half of the time series in a to construct a library of patterns, we use the simplex projection methods described in the text to predict one week ($T_p=1$) into the future from each point in the second half of the series ($N=415$); here $E=3$, and $\tau=1$. The correlation coefficient between predicted and observed values is $\rho=0.42$ ($P<10^{-4}$ for $N=415$); the best autoregressive linear predictions (composed of three optimal linear maps) give $\rho=0.13$, which is significantly less than the nonlinear result ($P<0.0005$). c, As in Figs 4c and 5c, ρ between predicted and observed values is shown as a function of the choice of E for predictions

two time steps into the future ($T_p=2$ and $\tau=1$). This figure indicates an optimal E of about 3, consistent with an attractor of dimension about 2. d, As in Figs 4d and 5d, ρ is shown as a function of the T_p (for $E=3$ and $\tau=1$). Here the correlation coefficient decreases with increasing prediction interval, in the manner characteristic of chaotic dynamics generated by a some low-dimensional attractor. That ρ is about 50% at best, however, indicates that roughly half the variance in the time series comes from additive noise. The dynamics of this system therefore seem to be intermediate between those of measles (for which Fig. 4d indicates deterministically chaotic dynamics) and chickenpox (for which Fig. 5d indicates purely additive noise superimposed on a seasonal cycle).

performs as well as our nonlinear predictor. We take this to indicate that chickenpox has a strong annual cycle (as does measles), with the fluctuations being additive noise (in contrast to measles, for which the fluctuations derive mainly from the dynamics).

The contrast between measles and chickenpox can be explained on biological grounds²¹. Measles has a fairly high 'basic reproductive rate' ($R_0 = 10\text{--}20$), and, after a brief interval of infectiousness, recovered individuals are immune and uninfected for life; these conditions tend to produce long-lasting 'inter-epidemic' oscillations, with a period of about 2 years, even in the simplest models²². This, in combination with seasonal patterns, makes it plausible that measles has complex dynamics. Chickenpox is less 'highly reproductive' (with R_0 values of about 8 to 10), and may recrudescence as shingles in later life; this makes for an infection less prone to show periodicities other than basic seasonal ones associated with schools opening and closing, and therefore indicates seasonal cycles with additive noise. Furthermore, reporting was compulsory for measles but not for chickenpox over the time period in question, which itself would be likely to make sampling error greater for chickenpox. Whatever the underlying biological explanation, the patterns in Figs 4d and 5d differ in much the same way as those illustrated in Fig. 2a for the artificially generated time series of Fig. 2, a and b. **Marine plankton.** A time series is provided by Allen's weekly record of marine planktonic diatoms gathered at Scripps Pier, San Diego, between 1920 and 1939 ($N = 830$). With the exception of the work of Tont²³, this collection of information has been little analysed, and not at all in the light of contemporary notions about nonlinear dynamics. The data comprise weekly totals of the numbers of individuals of all diatom species, tallied in daily seawater samples collected over ~20 years. As for our analysis of the measles and chickenpox data above, we do not 'smooth' the diatom data in any of the usual ways, although we take first-differences for reasons stated earlier. The resulting time series is shown in Fig. 6a.

The results of using the first half of the diatom series to predict the second half are shown in the usual way in Fig. 6b. Figure 6c shows ρ between predicted and observed results looking one time step ahead ($T_p = 1$), as a function of E . The optimum embedding dimension seems to be about 3. This value for E is consistent with our independent analysis of the data using the Grassberger-Procaccia algorithm, which indicates that $D \approx 2$.

Figure 6d shows ρ as a function of T_p (for $E = 3$). The consistent decay in predictive power as one extrapolates further into the future is consistent with the dynamics of the diatom population being partly governed by a chaotic attractor. This view is supported by the significantly better fit of the nonlinear predictor as compared with the optimal linear autoregressive model ($P < 0.0005$). We note, however, that deterministic chaos at best accounts for about 50% of the variance, with the rest presumably deriving from additive noise; the relatively low dimension of the attractor for diatoms compared with measles makes it plausible that the noisier fit of predicted weekly fluctuations in diatoms, versus the predicted monthly fluctuations in measles, reflects a much higher sampling variance for diatoms than for reported measles cases.

Conclusion

The forecasting technique discussed here is phenomenological in that it attempts to assess the qualitative character of a system's dynamics—and to make short-range predictions based on that understanding—without attempting to provide an understanding of the biological or physical mechanisms that ultimately govern the behaviour of the system. This often contrasts strongly with the laboratory and field-experiment approaches that are used to elucidate detailed mechanisms by, for example, many population biologists. The approach outlined here splits the time series into two parts, and makes inferences about the dynamical nature of the system by examining the way in which ρ (the correlation coefficient between predicted and observed results for the second part of the series) varies with prediction interval, T_p , and embedding dimension, E ; given the low densities of most time series in population biology, we share Ruelle's²⁸ lack of confidence in a direct assessment of the dimension of any putative attractor by Grassberger-Procaccia or other algorithms. Our approach works with artificially generated time series (for which we know the actual dynamics, and the underlying mechanisms, by definition), and it seems to give sensible answers with the observed time series for measles, chickenpox and diatoms (deterministic chaos in one case, seasonal cycles with additive noise in another, and a mixture of chaos and additive noise in the third). We hope to see the approach applied to other examples of noisy time series in population biology, and in other disciplines in which time series are typically sparse. □

Received 12 July 1989; accepted 19 February 1990.

1. Lorenz, E. N. *J. Atmos. Sci.* **26**, 636–646 (1969).
2. Tong, H. & Lim, K. S. *J. R. Statist. Soc. A* **42**, 245–292 (1980).
3. Farmer, J. D. & Sidorowich, J. J. *Phys. Rev. Lett.* **62**, 845–848 (1987).
4. Priestly, M. B. *J. Time Series Analysis* **1**, 47–71 (1980).
5. Edkman, J. P. & Ruell, D. *Rev. mod. Phys.* **57**, 617–619 (1985).
6. Crutchfield, J. P. & MacNamara, B. S. *Complex Systems* **1**, 417–452 (1987).
7. Casdagli, M. *Physica D* **35**, 335–356 (1989).
8. Grassberger, P. & Procaccia, I. *Phys. Rev. Lett.* **50**, 346–369 (1983).
9. Schwartz, I. *J. math. Biol.* **21**, 347–361 (1985).
10. Schaffer, W. M. & Kot, M. *J. theor. Biol.* **112**, 403–407 (1985).
11. Schaffer, W. M. & Kot, M. in *Chaos: An Introduction* (ed. Holden, A. V.) (Princeton Univ. Press, 1986).
12. Schaffer, W. M., Ellner, S. & Kot, M. *J. math. Biol.* **24**, 479–523 (1986).
13. Schaffer, W. M., Olsen, L. F., Truty, G. L., Fulmer, S. L. & Graser, D. J. in *From Chemical to Biological Organization* (eds Markus, M., Muller, S. C. & Nicolis, G.) (Springer-Verlag, New York, 1988).
14. Yule, G. U. *Phil. Trans. R. Soc. A* **226**, 267–278 (1927).
15. Takens, F. in *Dynamical Systems and Turbulence* (Springer-Verlag, Berlin, 1981).
16. Farmer, J. D. & Sidorowich, J. J. in *Evolution, Learning and Cognition* (ed. Lee, Y. C.) (World Scientific, New York, 1989).

17. London, W. P. & Yorke, J. A. *Am. J. Epidem.* **98**, 453 (1973).
18. Pimm, S. L. & Redfearn, A. *Nature* **334**, 613–614 (1988).
19. Lawton, J. H. *Nature* **334**, 563 (1988).
20. Helsenstein, U. *Statist. Med.* **5**, 37–47 (1986).
21. Anderson, R. M. & May, R. M. *Nature* **318**, 323–329 (1985).
22. Anderson, R. M., Grenfell, B. T. & May, R. M. *J. Hyg.* **93**, 587–608 (1984).
23. Tont, S. A. *J. mar. Res.* **39**, 191–201 (1981).
24. Varosi, F., Grebogi, C. & Yorke, J. A. *Phys. Lett. A* **124**, 59–64 (1987).
25. Abarbanel, H. D., Kadtke, J. B. & Brown, R. *Phys. Rev. B* **41**, 1782–1807 (1990).
26. Mees, A. I. Research Report No. 8 (Dept Mathematics, University of Western Australia, 1989).
27. Drepper, F. R. in *Erodynamics* (eds Wolff, W., Soeder, C. J. & Drepper, F. R.) (Springer-Verlag, New York, 1988).
28. Ruelle, D. *Proc. R. Soc. A* (in the press).

ACKNOWLEDGEMENTS. We thank Henry Abarbanel, Martin Casdagli, Sir David Cox, Doyne Farmer, Arnold Mandell, John Sidorowich and Bill Schaffer for helpful comments, John McGowan and Sargun Tont for directing us to Allen's phytoplankton records, Alan Trombla for computing assistance, and the United States NSF and the Royal Society for supporting this research.

$\beta 2$ -Microglobulin deficient mice lack CD4⁺8⁺ cytolytic T cells

Maarten Zijlstra, Mark Bix*, Neil E. Simister, Janet M. Loring, David H. Raulet* & Rudolf Jaenisch

Whitehead Institute for Biomedical Research, Nine Cambridge Center, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

* Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

Mice homozygous for a $\beta 2$ -microglobulin gene disruption do not express any detectable $\beta 2$ -m protein. They express little if any functional major histocompatibility complex (MHC) class I antigen on the cell surface yet are fertile and apparently healthy. They show a normal distribution of $\gamma\delta$, CD4⁺8⁺ and CD4⁺8⁺ T cells, but have no mature CD4⁺8⁺ T cells and are defective in CD4⁺8⁺ T cell-mediated cytotoxicity. Our results strongly support earlier evidence that MHC class I molecules are crucial for positive selection of T cell antigen receptor $\alpha\beta$ ⁺ CD4⁺8⁺ T cells in the thymus and call into question the non-immune functions that have been ascribed to MHC class I molecules.

$\beta 2$ -MICROGLOBULIN ($\beta 2$ -m) is a polypeptide of relative molecular mass 12,000 (12K), which is activated in mouse embryos by the 2-cell stage¹, and associates with the heavy chain of the polymorphic MHC class I proteins encoded by the H2-K, H2-L/D and Qa/Tla loci². The role of MHC class I proteins in the presentation of antigens to the immune system and in the development of the T-cell repertoire is well documented²⁻⁶, but it has been suggested that MHC molecules have other non-immunological functions, for example, as differentiation antigens^{7,8}, or in the function of hormone receptors⁹⁻¹³ or as olfactory cues influencing mating behaviour¹⁴⁻¹⁵. Furthermore, it has been demonstrated that $\beta 2$ -m associates with the Fc receptor in neonatal gut cells¹⁶, induces collagenase activity in fibroblasts¹⁷, and may serve as a chemotactic protein in the fetal thymus¹⁸. We and others have disrupted the $\beta 2$ -m gene by targeted mutation^{19,20}, and generated a mutant mouse strain that is unable to express $\beta 2$ -m protein¹⁹ and therefore most class I molecules, which require $\beta 2$ -m for assembly and cell surface expression.

Normal development

Mice heterozygous for the disrupted $\beta 2$ -m gene were intercrossed to derive animals homozygous for the mutation. DNA was isolated from embryos between day 14 and 18 of gestation, or from the tails of weanling mice, and the genotype of each animal was determined by Southern blot analysis as described¹⁹. Table 1 shows that 9 of 33 embryos and 23 of 101 adults were homozygous for the mutated $\beta 2$ -m gene. Homozygotes were indistinguishable from heterozygous or wild-type littermates, had normal body weight and, on autopsy, showed no noticeable alterations in any organ. When bred, normal sized litters were born and raised by homozygous parents. These results indicate that the mutation has no obviously detrimental effect on the well-being or breeding performance of the animals.

Truncated $\beta 2$ -m mRNA

Transcription of the wild-type (+/+) $\beta 2$ -m gene results in two messenger RNA species of 0.8 and 1.0 kilobases (kb) that are

due to the use of alternative polyadenylation signals in exon 4 (Fig. 1a, ref. 21). The mutant $\beta 2$ -m gene contains a 1.1-kb fragment of plasmid pMCneo inserted into exon 2 which is transcribed from the *tk* promoter. The inserted *neo* gene has the same transcriptional orientation as the disrupted $\beta 2$ -m gene and lacks a polyadenylation signal.

To characterize class I-specific mRNA in mutant mice, total as well as poly(A)⁺-selected RNA was isolated from a variety of tissues of +/+, +/- (heterozygous) and -/- (homozygous) mice or from embryonic fibroblasts and examined by northern blot analysis. The blots were hybridized to a $\beta 2$ -m probe, the *neo* probe and an MHC class I-heavy chain probe. Figure 1b shows two strong bands at 0.8 and 1.0 kb expected for $\beta 2$ -m mRNA in interferon- γ treated +/+ and +/- fibroblasts, whereas untreated cells showed a much weaker signal. In contrast, homozygous mutant cells did not synthesize the normal $\beta 2$ -m mRNA species, but showed instead a band at ~2.0 kb which was also seen in +/- cells, and is expected for an RNA species initiated at the $\beta 2$ -m promoter, transcribed through the *neo* cassette and terminated in exon 4. The intensity of the signal suggested that this RNA was much less abundant than the wild-type $\beta 2$ -m RNA. Similar hybridization signals were detected in liver, kidney, spleen, brain and lung RNA of adult mutant or wild-type mice (Fig. 1c, and data not shown). In addition, a faint signal of ~1.5 kb was seen in +/- and -/- animals, which may correspond to RNA initiated at the *tk* promoter.

We do not know why transcription from the $\beta 2$ -m promoter is impaired in cells of homozygous mice. It is possible that the *neo* insert exerts an inhibitory effect on $\beta 2$ -m promoter usage or that the read-through RNA is less stable. Alternatively, additional mutations not detectable by Southern analysis may have occurred in the promoter region and interfere with its function. Hybridization to the MHC class I heavy chain probe showed the expected signal of 1.6 kb with the same intensity in animals of all three genotypes (Fig. 1b, c). The results shown in Fig. 1 indicate that the $\beta 2$ -m gene disruption prevents synthesis of normal $\beta 2$ -m RNA but does not interfere with the transcription or stability of MHC class I-heavy chain RNA.

No $\beta 2$ -m protein

Embryonic fibroblasts were treated with interferon- γ , labelled with [³⁵S]methionine and protein extracts subjected to immune precipitation using $\beta 2$ -m and several MHC class I-specific antibodies. Figure 2a shows that the expected 12K protein was detected by precipitation with the $\beta 2$ -m specific antiserum in +/+ and +/- cells, but was absent in -/- cells. Recognition of the H-2K^b heavy chain by monoclonal antibody 5F1.1.24 (anti H-2K^b, $\alpha 2$ domain) was dependent on association of the heavy chain with $\beta 2$ -m because the expected 46K band was seen in wild-type and heterozygous cells but was not detected in mutant cells. In contrast, immunoprecipitation with monoclonal antibody 28-14-8S (anti H-2D^b, $\alpha 3$ domain) precipitated 44K and 46K heavy chain bands and the 12K $\beta 2$ -m band in wild-type and heterozygous cells, as well as a weaker 44K band in homozygous cells. The 44K molecule most probably represents an immature precursor molecule still containing the terminal high

mannose residues²². These results suggest incomplete or altered processing of the 44K H-2D^b polypeptide in the absence of β 2-m²². These results also confirm earlier studies that monoclonal 28-14-8S can recognize H-2D^b molecules even when not associated with β 2-m²³⁻²⁵.

MHC class I cell surface expression

Surface expression of MHC class I molecules was examined by incubating purified CD4⁺8⁺ T cells with a panel of MHC class I specific antibodies and evaluated by fluorescence-activated cell sorting (FACS) analyses. The data shown in Fig. 2b and Table 2 failed to reveal any staining of cells from homozygous

mice with any of the β 2-m, H-2K^b and Qa-2 specific antibodies. However, incubation with several different anti-D^b monoclonal antibodies resulted in detectable staining which was reduced 20-fold or more when compared with wild-type cells. This observation corroborates the notion that the H-2D^b molecule can reach the cell surface even in the absence of endogenous β 2-m²³⁻²⁵. The native configuration of D^b as detected by these antibodies may, however, depend on exogenous β 2-m derived from the culture medium. Results of a preliminary staining experiment performed in serum-free medium suggest that decreased D^b-specific staining occurs in the absence of exogenous β 2-m.

No functional Fc receptor

Recently, the β 2-m protein has been identified as the smaller component of the Fc receptor that mediates the uptake of IgG from milk in intestinal cells of neonatal rats¹⁶. To examine the role of β 2-m for the biological function of the mouse intestinal Fc receptor, brush borders were isolated from the small intestine of 11-day-old littermates and tested for the binding of [¹²⁵I]-labelled IgG²⁶. The results demonstrate that intestinal cells from homozygous mutant mice fail to show significant binding of IgG ($0.6\% \pm 1.8$ (mean $\pm$ s.d.) of total c.p.m. bound; $n = 4$). The amount of IgG bound to brush borders from heterozygotes ($6.2\% \pm 1.9$; $n = 4$) and wild-type mice (6.9% , 4.3% , $n = 2$) was indistinguishable. This indicates that the Fc receptor heavy chain in mouse must associate with β 2-m protein for functional expression on the cell surface.

No TCR $\alpha\beta^+$ CD4⁺8⁺ T Cells

Lymphoid organs derived from F2 animals of the three genotypes were characterized by two- and three-colour FACS analyses for the presence of different subsets of T cells. In both the adult thymus and spleen of 11-day-old adult homozygous mutant mice, a dramatic 100–150-fold reduction in T-cell antigen receptor (TCR) $\alpha\beta^+$ CD4⁺8⁺ T cells was observed (Fig. 3, and data not shown). No differences between heterozygote and wild-type cells were seen (data not shown). In the thymus of

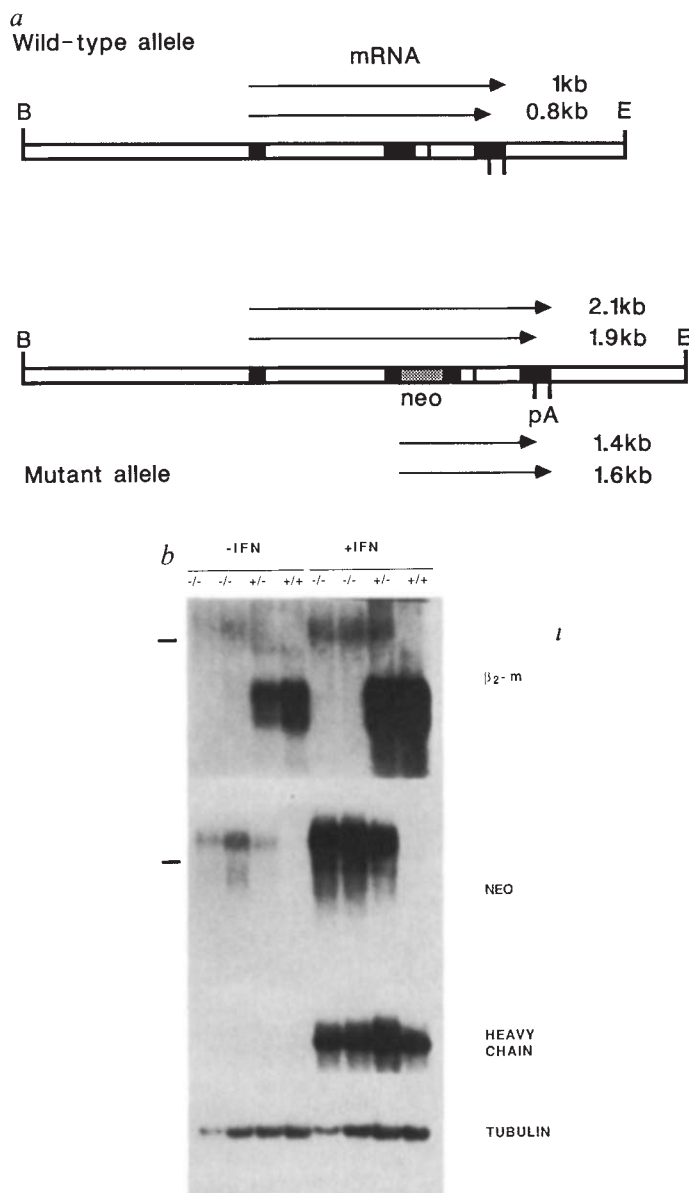
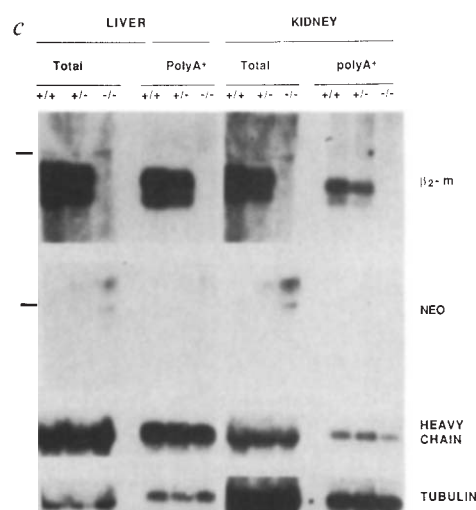


FIG. 1 Expression of β 2-m. *a*, Structure of the wild-type (wt) and mutant β 2-m genes and the predicted sizes of correctly spliced β 2-m and/or *neo*-specific mRNA transcripts. A 350-base pair (bp) *Pst*I-*Eco*RI fragment of β 2-m2 complementary DNA⁴⁷ was used as a β 2-m-specific hybridization probe. This fragment spans the 5' start site in exon 1 to the *Eco*RI site in exon 2. The following fragments were also used as hybridization probes: a 910-bp *Eco*RI-*Bam*HI fragment derived from pMCineo⁴⁸ (*neo* probe); a 1.4-kb *Pst*I fragment from pH2-d-37 containing a complete cDNA of H-2K^d gene⁴⁹ (MHC class I heavy chain probe); and a 1.6-kb *Pst*I fragment containing the entire rat α -tubulin cDNA (tubulin probe). *b*, Northern blot analysis of 15 μ g total cellular RNA from primary fibroblasts derived from 14-day-old F2



embryos of indicated genotype. Cells were grown in DMEM supplemented with 10% FCS, 1x non-essential amino-acids (Gibco). Interferon treatment was performed for 24 hours by supplementing the medium with 500 U ml⁻¹ recombinant mouse IFN- γ (Amgen biologicals). *c*, Northern blot analysis of 15 μ g total cellular RNA or 0.5 μ g oligo(dT)-selected poly(A)⁺ mRNA (one cycle) derived from adult liver and kidney of F2 animals of indicated genotypes. Total cellular RNA was isolated by the LiCl-urea method and separated on a 1.5% formaldehyde-treated agarose gel and blot-hybridized by standard procedures. Filters were stripped of hybridizing probes by treatment for 20 min in 10 mM Tris buffer, pH 7.5, 1% SDS, at 80 °C before reprobing. Bars indicate the migration of the 18S ribosomal RNA.

TABLE 1 Transmission of mutant $\beta 2$ -m gene

Parents	No. of Litters (age)	Genotype of progeny (no.)		
		-/-	+/-	+/+
+/- $\times$ +/-	4 (embryonic)	9	17	7
	21 (postnatal)	23	52	26
-/- $\times$ -/-	2 (postnatal)	17	0	0

F1 mice (129 $\times$ C57BL/6, both haplotype H-2^b) were intercrossed and offspring were genotyped as described¹⁹. -/-, Homozygous; +/-, heterozygous; +/+, wild type.

homozygous mutant mice the presence of the populations of TCR $\alpha\beta^{\text{dim}}$ CD4⁺8⁺ and $\alpha\beta^+$ CD4⁺8⁻ T cells was unaltered (TCR staining not shown). Also, $\alpha\beta^+$ CD4⁺8⁻ thymocytes were present in normal numbers in the thymus of young (day 11) mice. These cells are thought to represent an intermediate between $\alpha\beta^+$ CD4⁺8⁻ and $\alpha\beta^{\text{dim}}$ CD4⁺8⁺ cells^{27,28}. Therefore our results imply that MHC class I cell-surface expression is only essential for the development of the TCR $\alpha\beta^+$ CD4⁺8⁺ T cells. This indicates that differentiation of $\alpha\beta^+$ CD4⁺8⁺ T cells from $\alpha\beta^{\text{dim}}$ CD4⁺8⁺ thymocytes requires interaction with class I MHC molecules. Finally, it should be noted that the presence of the MHC class II-restricted CD4⁺8⁻ T cells (Fig. 3) and surface Ig⁺B cells (data not shown) are unchanged in the homozygous mutant mice.

Presence of γ/δ^+ T cells unaffected

There is evidence that at least a portion of TCR $\alpha\beta^-/\gamma\delta^+$ T cells may be restricted by MHC class I-like molecules encoded by the Qa/TL region of the H-2 complex²⁹⁻³². Therefore thymuses obtained from 17.5-day-old embryos and thymus and spleen from adult mice were examined for the presence of $\gamma\delta^+$ T cells. The number of $\gamma\delta$ cells was not significantly affected by the $\beta 2$ -m mutation (Fig. 3g, h). Furthermore, the abundance in the fetal (E17.5) thymus of a subset of $\gamma\delta$ cells expressing the V $_{\gamma 3}$

TABLE 2 Cell surface expression of MHC class I molecules

1st Reagent	Specificity	Mean fluorescence intensity			Ref.
		+/+ F2	-/- F2	B10.BR	
none	—	1.7	1.5	1.6	
NEI-026	$\beta 2$ -m ^b	168.2	1.5	73.4	
M1/42.3.9.8	K ^b , H-2 ^k	347.7	2.3*	154.7	39
H141-11	K ^b , K ^k	219.4	2.1	128.3	40
K10-56.1	K ^b ($\alpha 1/\alpha 2$)	8.2	1.5	1.3	41
K7-309	K ^b ($\alpha 1/\alpha 2$)	6.7	1.6	1.3	41
B8-24-3	K ^b ($\alpha 1$)	113.6	1.9	1.3	42
28-13-3S	K ^b ($\alpha 2$)	139.5	2.0	1.3	43
5F1.1.24	K ^b ($\alpha 2$)	61.2	1.7	3.1	44
28-14-8S	D ^b ($\alpha 3$)	119.3	6.3	1.3	43
H141-31-10	D ^b ($\alpha 2$)	36.0	1.7	1.3	40
B22-249R1	D ^b ($\alpha 1$) priv.	146.9	8.5	1.3	40
27-11-13S	D ^b ($\alpha 1$)	112.0	3.5	1.3	43
D3-262	Qa-2($\alpha 3$)	38.6	1.6	1.3	45
1-5-9	Qa-2($\alpha 1/\alpha 2$)	12.1	1.6	1.3	46
1-4-4	Qa-2($\alpha 1/\alpha 2$)	29.6	1.6	1.3	46

CD4⁺ T cells were purified from lymph node cells as described in Fig. 2, reacted with the indicated monoclonal antibody (Mab) followed by FITC-goat anti-mouse IgG + M (or FITC-goat anti-rat IgG for M1/42.3.9.8), and 1×10^4 stained cells analyzed on an Epics C flow cytometer. The domain specificity of the Mab, where known, is indicated. T cell purification, staining and analysis were in the presence of 5% FCS. The numbers refer to mean linear fluorescence intensity.

* Control mean fluorescence intensity with FITC-goat anti-rat IgG is 2.0.

gene segment was normal in the $\beta 2$ -m mutant mice (see legend, Fig. 3).

No cytotoxic T cells

Murine alloreactive cytotoxic T lymphocytes (CTL) are almost exclusively TCR $\alpha\beta^+$ CD4⁺8⁺ (refs 2-6). Therefore spleen cells derived from F2 animals were tested for the presence of CTL-precursors (CTL-p) in a bulk-mixed lymphocyte culture (MLC)

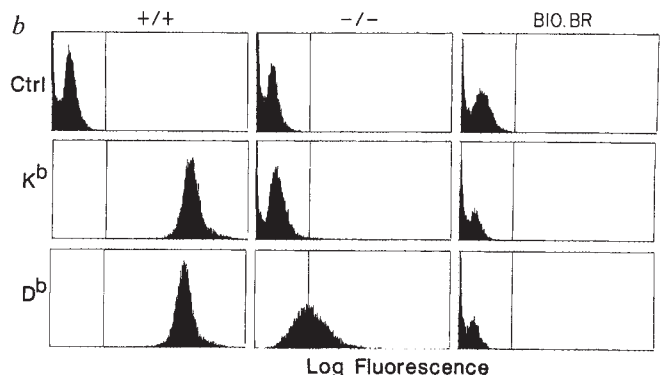
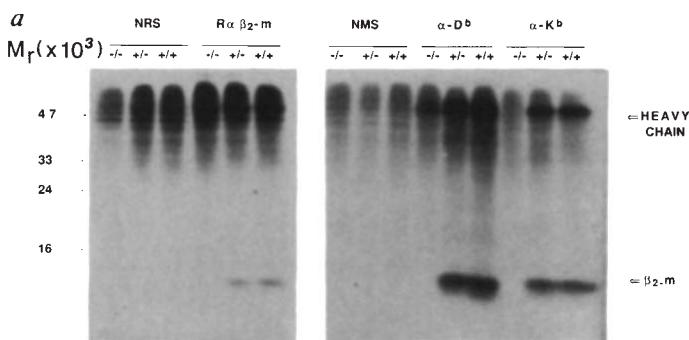


FIG. 2 Expression of MHC class I molecules. *a*, Immunoprecipitations of MHC class I molecules from metabolically labelled embryonic fibroblasts. Fibroblasts derived from day-14 embryos were routinely cultured (see Fig. 1b). Normal medium was supplemented with 500 U ml⁻¹ IFN- γ 24 h before labelling. Subsequently the nearly confluent cells were washed once with medium containing 9 volumes methionine-free DMEM to 1 volume normal DMEM, supplemented with 5% dialysed FCS (label medium). Thereafter cells were incubated for 2.5 h at 37 °C in label medium with the addition of 0.2 mCi ml⁻¹ [³⁵S]-L-methionine (Amersham). Subsequently cultures were washed once with normal medium and chased for 1 h at 37 °C in normal medium. Thereafter cultures were washed once with ice-cold PBS buffer and subsequently lysed in ice-cold 100 mM Tris, pH 7.4, 50 mM NaCl, 1 mM MgCl₂, 0.5% Nonidet P-40 and 1% aprotinin. Lysates were centrifuged at 12,000g (15 min) and 8×10^6 trichloroacetic acid precipitable c.p.m. of the supernatant were used for subsequent immunoprecipitation of class I molecules. Immunoprecipitation was performed as described by Williams *et al.*²⁴ with the use of protein A-Sepharose (Pharmacia). Eluted antigen was subjected to SDS-PAGE analysis using a 12% Laemmli gel⁵⁰. Gels were fixed in acetic

acid, incubated with 22% PPO (w/v) in acetic acid, dried and exposed to preflashed Kodak X-OAR5 film at -70 °C. Bars indicate the migration of prestained protein markers (BioRad). NRS, normal rabbit serum; anti- $\beta 2$ -m, rabbit anti-human $\beta 2$ -m (Serotec); NMS, normal mouse serum; anti-D^b, monoclonal antibody 28-14-8S ($\alpha 3$ domain); anti-K^b, monoclonal antibody 5F1.1.24 ($\alpha 2$ domain). *b*, Cell surface expression of MHC class I molecules of purified T cells obtained from lymph nodes of 4-week-old wild-type (+/+) and homozygous mutant (-/-) F2 animals. T cells from 5-week-old H-2^k (B10.BR) mice served as negative controls for staining. Purified CD4⁺ T cells were prepared by passage of lymph node cells over nylon wool columns and panning the nonadherent cells on plates coated with anti-CD4 (GK1.5) antibodies as described⁵¹. Cells were reacted in the first stage with reagents specific for the $\alpha 2$ domain of K^b (28-13-3S), the $\alpha 3$ domain of D^b (28-14-8S), or no antibody, followed in the second stage with FITC-conjugated goat anti-mouse IgG + M (Kirkegaard and Perry, Gaithersburg, Maryland). Ten thousand cells were analysed on an EPICS C flow cytometer equipped with a quartz flow cell.

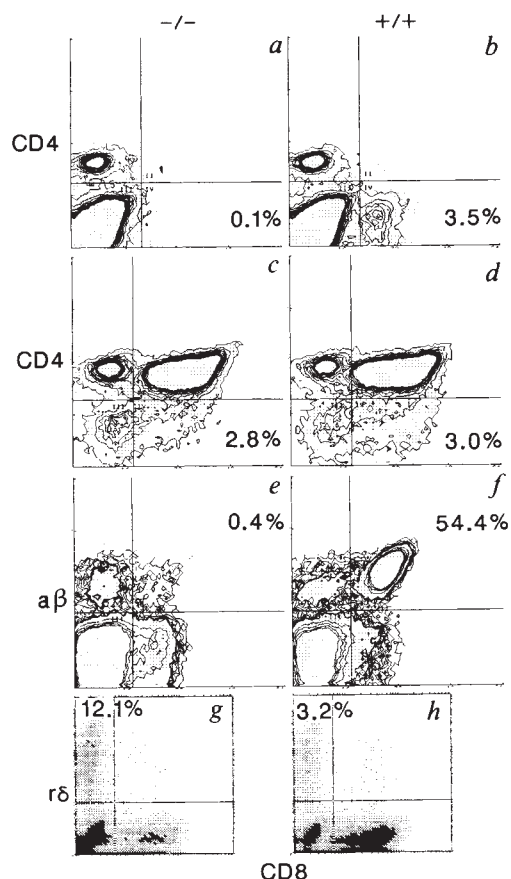


FIG. 3 Analysis of T-cell subsets in wild-type (+/+) and homozygous mutant (-/-) mice. Splenocytes (a, b) and thymocytes (c, d) from 11-day-old mice were stained with anti-CD4-phycoerythrin and anti-CD8-FITC (Becton and Dickinson, Mountainview, California). In panels e-h, CD8⁺CD4⁻ cells and γδ cells were first enriched from thymocyte populations of 4–5-week-old mice by eliminating CD4⁺ cells (most γδ cells are CD4⁻CD8⁺). The enriched cells were stained with anti-αβ TCR monoclonal antibody (H57-597)⁵² (e, f), or anti-γδ TCR monoclonal antibody (UC7-13D5, J. Bluestone, unpublished data) (g, h) (ordinate), and with anti-CD8 antibody (e-h) (abscissa). The percentages of the cells in relevant quadrants are indicated on the figure. In e and g, the lack of CD8⁺ cells leads to a twofold enrichment of γδ cells; in several experiments there was no obvious difference between +/+ and -/- mice in the frequency of CD4⁻CD8⁺ cells that are γδ positive. In addition, the presence of γδ cells was examined in unfractionated thymocytes of day-17.5 embryos. No significant differences were seen in the number of cells reactive with anti-γδ antibody UC7-13D5: -/-, 1.86% ± 0.27 (n=6); +/-, 1.99% ± 0.35 (n=13) and +/+, 2.40% ± 0.57 (n=5). Also, no changes in the number of cells reactive with antibody F-536 (anti-V_{γ3})⁵³ were noted: -/-, 0.88% ± 0.22 (n=6); +/-, 0.80% ± 0.23 (n=13) and +/+, 1.10% ± 0.24 (n=5).

METHODS. To eliminate CD4⁺ cells, thymocytes were incubated with anti-CD4 (GK1.5) monoclonal antibody and complement for 40 min. Viable cells were purified on Ficol-Isoopaque gradients and subjected to a second round of killing with mouse anti-rat α light chain (MAR18.5) antibody plus complement to eliminate residual CD4⁺ cells with bound GK1.5. The viable cells were again purified on Ficol-Isoopaque gradients. To stain αβ TCR⁺ versus CD8⁺, the enriched cells were reacted with H57-597-biotin followed in a second step with allophycocyanin-streptavidin (Becton Dickinson) and anti-CD8-FITC. To stain γδ TCR versus CD8, enriched cells were first reacted with UC7-13D5 culture supernatant, followed in a second step by goat-anti-hamster IgG-phycoerythrin (reagent adsorbed with rat and mouse IgG; Caltag, California). After washing, the cells were incubated with rat-IgG to ensure that there were no free rat immunoglobulin-binding sites, and subsequently reacted with anti-CD8-FITC. In all cases the negative controls using all reagents except the TCR-specific first reagents, gave insignificant numbers of positive cells. Cursors were set based on the negative control samples. Dead cells were excluded based on forward and 90-degree light scatter characteristics. One hundred thousand cells were analysed on a FACSTAR+ (a-f) or 3 × 10⁴ cells on an EPICS C (g, h) flow cytometer.

against completely allogeneic BALB/c (H-2^d) cells. Spleen cells derived from homozygous mutants were virtually devoid of any CTL-p (Fig. 4c). The CTL responses of heterozygous animals were indistinguishable from those obtained with wild-type animals (data not shown). In addition, we examined whether the mutant cells could stimulate CTL-p or serve as target cells for established CTL. As shown in Fig. 4a, β2-m-negative spleen cells fail to elicit a significant CTL response by BALB/c responders. Similar results were obtained with B10.BR responders (data not shown). Therefore the low residual H-2D^b cells surface expression in homozygous mutant cells is clearly not sufficient to trigger a vigorous CTL response.

Finally we find that mutant cells can serve as target cells for anti-H-2^b CTL generated in conventional MLC, although about ninefold more CTL are required to lyse mutant compared with wild-type targets (Fig. 4b). This residual killing can be accounted for by three hypotheses. First, it is possible that D^b molecules assume a functional conformation even in the absence of β2-m, albeit at a dramatically reduced level. Alternatively, bovine β2-m from the serum-containing medium may associate with cell surface D^b and facilitate refolding of the molecule. This latter hypothesis is consistent with published data demonstrating the binding of serum β2-m to class I molecules on cultured cells³³. Finally, it is possible that the residual CTL activity is specific for non-class I antigens, although no lysis of B10.BR targets was observed (Fig. 4b) and highly purified mutant T-cell blasts

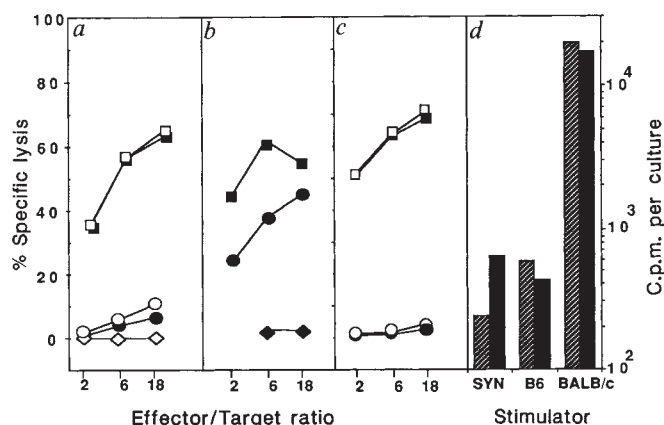


FIG. 4 Functional studies of β2-m mutant mice. a, β2-m mutant (-/-) F2 spleen cells (○, ●) from two animals stimulate little or no BALB/c (H-2^d) cytotoxic T lymphocytes (CTL) reactive with H-2^b. Shown for comparison are cultures stimulated with wild-type littermate (+/+) (□, ■) and syngeneic spleen cells (◇). All cultures were tested against C57BL/6 (H-2^b) target cells. b, β2-m mutant (-/-) concavalin A-induced blast cells (●) serve as targets for conventional anti-H-2^b CTL. The CTL were from secondary mixed lymphocyte culture of B10.BR spleen cells stimulated with wild-type (+/+) F2 spleen cells. Shown for comparison is the lysis of wild-type littermate target cells (■) and B10.BR target cells (◆). Approximately nine times more of the CTL are required to achieve the same level of lysis of β2-m^{-/-} targets as of the β2-m^{+/+} targets. c, Spleen cells of two β2-m mutant (-/-) mice (○, ●) fail to generate allo-specific CTL when stimulated with BALB/c (H-2^d) spleen cells. Shown for comparison are the responses of two wild-type littermates (□, ■). Cultures were tested on BALB/c concavalin A-induced target cells. d, β2-m mutant (-/-) lymph node cells proliferate in response to allogeneic (BALB/c, H-2^d) stimulator cells (hatched bars). The response of a wild-type littermate is shown for comparison (solid bars). Cultures stimulated with syngeneic or C57BL/6 spleen cells served as negative controls. Proliferative response was determined between days 5–6 of culture by pulsing overnight with [³H]thymidine (0.5 μCi).

METHODS. Details of mixed lymphocyte culture and CTL assay procedures are as described⁵⁴. In all cases target cells were from cultures of spleen cells incubated for 2 days with 2 μg ml⁻¹ concanavalin A, and labelled for 1 h with 200 μCi ⁵¹Cr. CTL assay was for 4 h. MLR (d) was performed in round-bottom microtitre wells with 1 × 10⁵ responder lymph node cells and 5 × 10⁵ irradiated (3,000 rads) stimulator spleen cells.

are lysed, arguing that the activity is not directed against class II molecules (data not shown).

When mutant cells were used as either responders (Fig. 4d) or as stimulator cells (data not shown) in a mixed lymphocyte reaction (MLR) the proliferative responses were similar to those with wild-type cells. This is consistent with the fact that the proliferation measured in an MLR is predominantly determined by the recognition of foreign MHC class II molecules by CD4⁺8⁺ T cells²⁻⁶.

Discussion

Mice homozygous for the $\beta 2\text{-m}$ gene disruption develop normally in the absence of detectable $\beta 2\text{-m}$ protein. Our results therefore question the widely held view that class I molecules have an essential role in the embryonic development of vertebrates, although the possibility that some class I molecules may be functional in the absence of $\beta 2\text{-m}$ has not been ruled out. That mutant embryos have a normal sized thymus argues that the function of $\beta 2\text{-m}$ as a chemotactic protein (thymotaxin) is not essential for thymus development.

Of interest is the apparently normal development of $\gamma\delta$ T cells in $\beta 2\text{-m}$ mutant mice. The physiological functions and specificity of $\gamma\delta$ cells are not yet understood, but it has been suggested that $\gamma\delta$ cells functionally interact with class I molecules, particularly those encoded by Qa/TL region genes telomeric to the H-2 complex. If $\gamma\delta$ cells naturally interact with class I molecules, it might be expected that like CD8⁺ T cells, interactions with class I molecules would be required for their differentiation. Although our data argue against this, it is still possible that only mature $\gamma\delta$ cells recognize class I molecules or antigens associated with class I molecules. In addition, it is possible that differentiation of only a subset of $\gamma\delta$ cells requires class I MHC molecules.

The most striking phenotype of the $\beta 2\text{-m}$ mutant mice is the

virtually complete absence of the mature $\alpha\beta^+$ CD4⁺8⁺ T cell subset, in both the thymus and peripheral lymphoid organs. These data argue that encounters with class I MHC molecules are essential for the differentiation of the $\alpha\beta^+$ CD4⁺8⁺ T cell subset, consistent with the conclusions drawn from studies of transgenic mice expressing a defined $\alpha\beta$ TCR in all their T cells³⁴⁻³⁶, as well as mice treated from birth with anti-class I MHC antibodies^{37,38}. Presumably, interaction with class I molecules induces CD4⁺8⁺ T cells to differentiate into CD4⁺8⁺ $\alpha\beta^+$ T cells, or rescues newly differentiated $\alpha\beta^+$ CD4⁺8⁺ T cells from programmed cell death. Conversely, our results indicate that differentiation of the other defined thymocyte subsets does not require interactions with class I molecules. Included in the latter category are CD4⁺8⁺ $\alpha\beta^{\text{dim}}$ thymocytes, CD4⁺8⁺ mature T cells, and CD4⁺8⁺ $\alpha\beta^-$ thymocytes, the latter thought to represent an intermediate between CD4⁺8⁺ thymocytes and CD4⁺8⁺ thymocytes^{27,28}.

As expected from the absence of mature CD4⁺8⁺ T cells, the CTL responses in homozygous mutant mice were abolished. The animals, when kept under pathogen-free conditions, appear to be healthy. It will, however, be interesting to study the role of MHC class I antigens in the response to viral or other infections or in tumorigenesis by exposing the animals to infectious agents or to carcinogens. Also, because of the deficiency in class I-dependent functions, homozygous mutant mice may serve as donors or recipients in transplantation experiments across histocompatibility barriers without treating the recipients with immunosuppressive regimens. Finally, transplantations of mutant bone marrow cells to normal mice should test the hypothesis that rejection of bone marrow (haemopoietic histocompatibility) can depend on the failure of marrow cells to express self class I molecules⁵⁵. Such studies will be greatly aided by the availability of a homozygous breeding colony. □

Received 20 February; accepted 12 March 1990.

1. Sawicki, J. A., Magnuson, T. & Epstein, C. J. *Nature* **294**, 450-451 (1981).
2. Klein, J. *et al.* *Nature* **291**, 455-460 (1981).
3. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51-117 (1979).
4. Bevan, M. J. & Fink, P. *Immunol. rev.* **42**, 3-19 (1978).
5. Von Boehmer, H. A. *Rev. Immun.* **6**, 309-326 (1988).
6. Marrack, P. & Kappler, J. *Immun. Today* **9**, 308-315 (1988).
7. Curtis, A. S. G. & Rooney, P. *Nature* **281**, 222-223 (1979).
8. Bartlett, P. F. & Edidin, M. *J. Cell Biol.* **77**, 377-388 (1978).
9. Verland, S. *et al.* *J. Immun.* **143**, 945-951 (1989).
10. Kittur, D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 1351-1355 (1987).
11. Hansen, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 3123-3126 (1989).
12. Schreiber, A. B., Schlessinger, J. & Edidin, M. *J. cell Biol.* **98**, 725-731 (1984).
13. Solano, A. R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 5087-5091 (1988).
14. Singh, P. B., Brown, R. E. & Roser, B. *Nature* **327**, 161-164 (1987).
15. Yamazaki, K. *et al.* *Science* **240**, 1331-1332 (1988).
16. Simister, N. E. & Mostov, K. E. *Nature* **337**, 184-187 (1989).
17. Brinckerhoff, C. E. *et al.* *Science* **243**, 655-657 (1989).
18. Dargemont, C. *et al.* *Science* **246**, 803-806 (1989).
19. Zijlstra, M. *et al.* *Nature* **342**, 435-438 (1989).
20. Koller, B. H. & Smithies, O. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8932-8935 (1989).
21. Parnes, J. R., Robinson, R. R. & Seidman, J. G. *Nature* **302**, 449-452 (1983).
22. Hansen, T. H., Myers, N. B. & Lee, D. R. *J. Immun.* **140**, 3522-3527 (1988).
23. Allen, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 7447-7451 (1987).
24. Williams, D. B. *et al.* *J. Immun.* **142**, 2796-2806 (1989).
25. Potter, T. A., Boyer, C., Schmitt-Verhulst, A. M., Golstein, P. & Rajan, T. V. *J. exp. Med.* **160**, 317-322 (1984).
26. Simister, N. E. & Rees, A. R. *Eur. J. Immun.* **15**, 733-738 (1985).
27. Nikolic-Zugic, J., Moore, M. W. & Bevan, M. J. *Eur. J. Immun.* **19**, 649-653 (1989).
28. Guidos, C. J., Weissman, I. L. & Adkins, B. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7542-7546 (1989).
29. Matis, L. A., Cron, R. & Bluestone, J. A. *Nature* **330**, 263-264 (1987).
30. Bonneville, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 5928-5932 (1989).
31. Vidovic, D. *et al.* *Nature* **340**, 646-650 (1989).
32. Raulet, D. H. A. *Rev. Immun.* **7**, 175-207 (1989).

33. Bernabeu, C. *et al.* *Nature* **308**, 642-645 (1985).
34. Kisielow, P. *et al.* *Nature* **333**, 742-747 (1988).
35. Von Boehmer, H., Teh, H. S. & Kisielow, P. *Immun. Today* **10**, 57-61 (1989).
36. Sha, W. C. *et al.* *Nature* **335**, 271-274 (1988).
37. Marusic-Galesic, S. *et al.* *Nature* **333**, 180-183 (1988).
38. Zuniga-Pflucker, J. C., Longo, D. L. & Kruisbeek, A. M. *Nature* **338**, 76-78 (1989).
39. Stallcup, K. C., Springer, T. A. & Mescher, M. F. *J. Immun.* **127**, 923-930 (1981).
40. Hämmerling, G., Rusch, E., Tada, N., Kimura, S. & Hämmerling, U. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4737-4741 (1982).
41. Lemke, H., Hämmerling, G. J. & Hämmerling, U. *Immun. Rev.* **47**, 175-206 (1979).
42. Kohler, G., Fischer-Lindahl, K. & Heusser, C. in *The Immune System* (eds Steinberg, S. G. & Lefkovits, L.) **2**, 202 (Karger, Basel, 1981).
43. Ozato, K. & Sachs, D. H. *J. Immun.* **126**, 317-321 (1981).
44. Sherman, L. A. & Randolph, C. P. *Immunogenetics* **12**, 183-186 (1981).
45. Lynes, M. A., Tonkonogy, S. & Flaherty, L. *J. Immun.* **129**, 928-934 (1982).
46. Sharrow, S. O., Arn, J. S., Stroynowski, I., Hood, L. & Sachs, D. H. *J. Immun.* **142**, 3495-3502 (1989).
47. Daniel, D. *et al.* *EMBO J.* **2**, 1061-1065 (1983).
48. Thomas, K. R. & Capecchi, M. R. *Cell* **51**, 503-512 (1987).
49. Lalanne, F. *et al.* *Nucleic Acids Res.* **11**, 1567-1577 (1983).
50. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
51. Holsti, M. A. & Raulet, D. H. *J. Immun.* **143**, 2514-2519 (1989).
52. Kubo, R. T., Born, W., Kappler, J. W., Marrack, P. & Pigeon, M. J. *Immun.* **142**, 2736-2742 (1989).
53. Havran, W. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 4185-4189 (1989).
54. Raulet, D. H. *et al.* *J. Immun.* **125**, 1136-1143 (1980).
55. Ohlen, C. *et al.* *Science* **246**, 666-668 (1989).

ACKNOWLEDGEMENTS. We thank Drs Jeffrey Bluestone, James P. Allison, Floris Shoe, Stanley Nathenson and Peter Walden for gifts of monoclonal antibodies and Drs Alan Korman, Kees Murre and William Strauss for critical reading of the manuscript. We also thank Dr Nan-Shih Liao for help and advice, Rob West for technical assistance, and Stewart Conner for expert assistance with flow cytometry. This work was supported by Cancer Research Institute fellowships to M.Z. (Fan Fox & Leslie R. Samuels Foundation Fellowship) and to N.E.S. (F.M. Kirby Foundation Fellowship), and a Frances and Edwin L. Cummings/CRI Investigator Award to D.H.R. We acknowledge the NIH for financial support (D.H.R. and R.J.).

Measuring black hole mass through variable line profiles from accretion disks

L. Stella*

EXOSAT Observatory, Astrophysics Division, Space Science Department of ESA, ESTEC, Keplerlaan 1, 2200 AG Noordwijk, The Netherlands, and Osservatorio Astronomico di Brera, Via Brera 28, 20121 Milano, Italy

* Affiliated to the International Center for Relativistic Astrophysics

THE emission lines in the optical spectra of several active galactic nuclei (AGNs) show a broad double-horned profile which is likely to be produced by matter orbiting the central massive black hole in a disk-like geometry¹⁻³. The unresolved and, sometimes, broad and redshifted Fe K α lines observed in the X-ray spectra of Seyfert galaxies⁴⁻⁶ and bright X-ray binaries⁷⁻¹⁰ might also be dominated by Doppler, transverse and gravitational shifts in a keplerian disk of tens to hundreds of Schwarzschild radii ($r_s = 2GM/c^2$) around the collapsed object^{11,12}. Line emission from these disks is most likely to be driven by photoionization or Compton heating by radiation from closer to the central object; flux variations in this radiation will cause changes in the disk line emissivity. Here we present detailed calculations of the response of a line from a relativistic keplerian disk to flux variations in the central regions. Two characteristic features form on opposite sides of the line, which, owing to light travel-time effects, gradually drift from the line wings, corresponding to the innermost disk regions, to the blue and red horns. Observations of these line-profile changes

could provide the first direct measurement of the mass of the central black hole in AGNs^{2,12,13}.

The Balmer line profiles of several radio galaxies show a clear double-horned structure with wings similar to, but much broader than, those observed from the disks of cataclysmic variables and spiral galaxies. Using a relativistic accretion disk to model the H α profile of Arp 102B (refs 2, 3) and H β profile of 3C 390.3 (ref. 1) provides estimates of the disk parameters, such as the inner and outer radii and the inclination of the line-emitting region.

Broad Fe K α emission lines at $\sim 6-7$ keV are observed in a number of bright X-ray binaries containing a black hole^{7,8} or a weakly magnetic neutron star^{9,10}. Owing to the low resolution of current observations ($E/\Delta E \leq 10$), the detailed line profiles are still unknown. Fe K α lines broader than 1 keV and significantly redshifted are seen in the black-hole candidates Cyg X-1 (ref. 7) and 4U1543-47 (ref. 8). Neither blending of different lines from different ionization stages of iron nor Compton broadening is likely to produce line widths ≥ 0.5 keV (ref. 11). Therefore it has been suggested that the line profile and centroid energy of Fe K-shell lines in bright X-ray binaries are determined primarily by bulk plasma motions in a keplerian disk irradiated with hard X-rays emitted close to the central collapsed object^{11,12}.

Direct evidence for broadening in the Fe K α lines observed from Seyfert I galaxies and quasars is still lacking. The variable XUV (X-ray/ultraviolet) excess in these systems suggests the presence of cold matter orbiting the massive black hole in a disk. Irradiation by hard X-rays can generate X-ray lines through fluorescence by cold material in the inner disk regions^{4-6,12}.

The estimated size of the line-emitting disk region ranges from

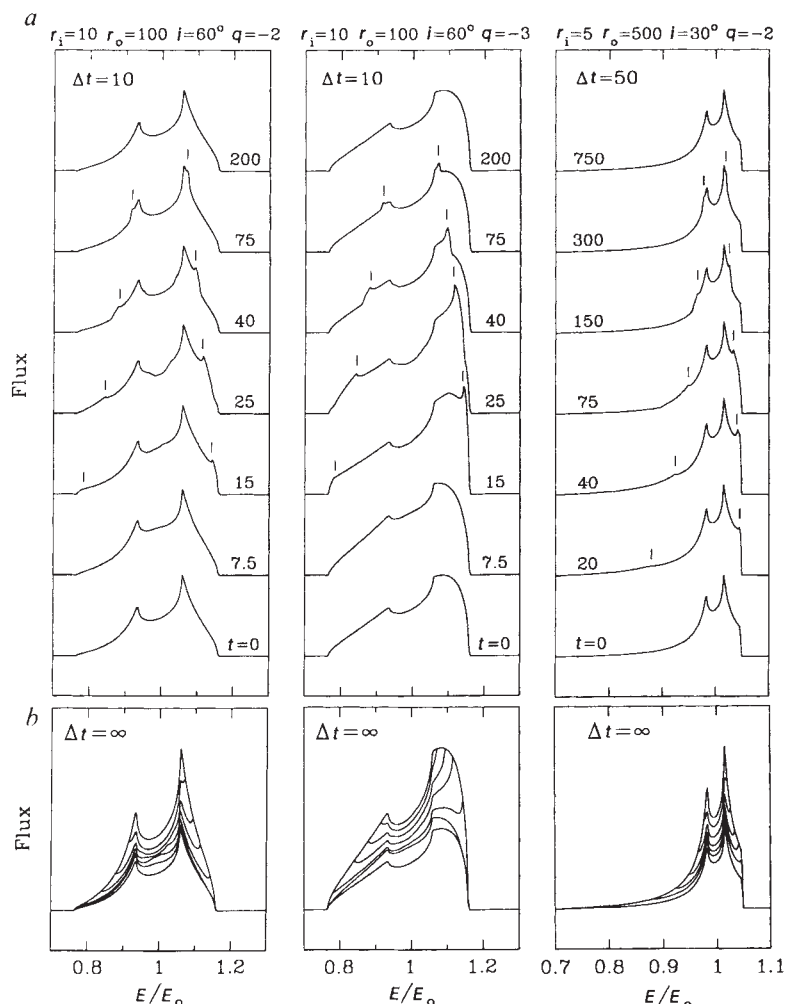


FIG. 1. Response of the line profile from a relativistic keplerian disk to flux variations in the central source, as a function of energy. Three examples are shown, corresponding to different choices of the parameters r_i , r_o , i and q . Radii are in the range of values inferred from the widths of Fe K α lines in accreting X-ray binaries; they are given in units of $r_s = 2.95 \times 10^5 (M/M_\odot)$ cm. A rectangular doubling of the flux from the central source, starting at $t_0 = 0$ and lasting for Δt , is assumed. The profiles are shown for seven different times t after the flare onset. All times refer to the observer and are expressed in units of the Schwarzschild radius light-crossing time $r_s/c = 0.98 \times 10^{-5} (M/M_\odot)$ s. The upper panels (a) show the response to short duration flares. The line profile response to very long flares is shown in the lower panels (b); times t are the same as in the upper panels. Here and in Fig. 2 the vertical lines represent the predicted position of the drifting features (see text).

tens to hundreds of times r_s for all these classes of accreting collapsed objects. Flux variations from the central source will induce changes in the disk line emissivity. Light travel-time effects cause the innermost line-emitting regions of the disk to respond first, followed by the outer regions¹². To give a quantitative description of this phenomenon we consider the case in which the dimensions of the central source of radiation are much smaller than the inner radius of the line-emitting portion of the disk. We assume a keplerian thin disk orbiting a Schwarzschild black hole. For an observer at infinity any variation in the flux from the central source will be 'echoed' by different locations in the disk at different times. In the limit of a point-like central source, neglecting all general relativistic delays in light propagation, a flux variation from the central source reaching a distant observer at a time t_0 will be echoed after a time $t - t_0$ by the disk region defined by

$$t - t_0 = r(1 - \sin i \sin \phi)/c \quad (1)$$

where i is the inclination angle from which the disk is observed and r and ϕ are the coordinates in the disk plane; $\phi = 0^\circ$ at the line of nodes and $0^\circ < \phi < 180^\circ$ in the half disk nearer the observer. The locus of constant t on the disk (the 'echo front') is an ellipse of eccentricity $\sin i$, with the most distant focus at $r = 0$.

The line profile at the observer's time t is determined by the inner radius r_i , the outer radius r_o , the inclination i , and the dependence of the disk emissivity, ϵ , on r and ϕ (I adopt the fully relativistic approach described in ref. 12). The fraction of the radiation from the central source intercepted by the disk at different radii and, therefore, the line emissivity depends on the detailed geometry of the central source and disk surface, and on a number of local physical conditions. It is customary to

assume that, in stationary conditions, the line emissivity resulting from the combination of all these effects can be described by a power law $\epsilon \propto r^q$. It is expected that $q \sim -2$ for radii of up to several times the dimensions of the central source and $q \sim -3$ beyond^{3,12}. If flux variations of the central source are not accompanied by substantial changes in the geometry or physical parameters of the disk, only the normalization of the line emissivity will vary in response. In particular, disk regions at different r and ϕ will respond with a time lag given by equation (1) for a distant observer at an inclination i . The relevant integrations can be carried out by specifying a model for the variability of the central source. For the sake of clarity we consider here only step-like variations; more general cases will be presented elsewhere.

Figure 1a shows three examples of line profiles from a keplerian accretion disk and their response to a step-like doubling of the flux from the central object, which lasts for a short time Δt . Radii and times are expressed in units of the Schwarzschild radius r_s and the corresponding light crossing time r_s/c , respectively, so that all figures can easily be rescaled for any mass. In general the profiles show a faster response in the line wings, after which the changes gradually move towards the core as the echo front propagates outwards. It should be noted, however, that shortly after the flare onset there is a temporary increase around the line centre between the blue and red horns. This occurs because the echo reaches first the line-emitting region around $\phi \approx 90^\circ$, where the components of the keplerian velocities toward the observer and the corresponding shifts are small. The effect is more pronounced for high inclinations (see first two panels of Fig. 1a for $t - t_0 = 7.5$ and $15 r_s/c$). As time increases the entire echo front is included within the line-emitting disk and propagates according to equation (1). In this case the profile generated by a thin annulus around the echo front has two sharp spikes, originating from the two zones close to the line of nodes where blueshift and redshift are highest. These sharp spikes generate the two drifting features on the wings of the line profile that are visible in Fig. 1. As time increases, the echo front propagates outwards through the disk and the two features drift towards the blue and red horns of the line profile as the keplerian velocities decrease. The slower drift of the blue feature is due to the fact that the decreasing Doppler blueshift is partly cancelled by the decreasing Doppler transverse and gravitational redshifts. Once the front echoing the final part of the source flare has propagated beyond the outer disk radius, the line will have returned to its original profile.

The position of the peak of the drifting features at a time $t - t_0$ is determined by the outermost echo front within the line-emitting portion of the disk, that is, the disk region 'echoing' the onset of the source flare. This is best seen by letting the flare duration increase to very large values. Figure 1b shows the evolution of the same profiles of Fig. 1a in response to a step-like flux doubling of the central source that lasts indefinitely. In this case the whole profile gradually swells as a larger and larger fraction of the disk doubles its line emissivity in response. Two very clear features still form on the blue and red wings of the line and drift inwards as time increases. A comparison of Fig. 1a and 1b shows that the positions of the red and blue drifting features are unchanged and therefore independent of the duration of the rectangular flare. At large times from the flare onset, the entire line-emitting disk is reached by the flux increase, and the line profile regains its original shape at an intensity a factor of two higher.

The drifting blue and red features provide a sensitive indicator of the inner disk regions. To exploit the information contained in them it is necessary to express their shift as a function of the disk parameters and time. This requires determining the locations on the echo front from which the most redshifted and blueshifted line signals are emitted. To this end we inserted equation (1) in equation (A2) of ref. 12 expressing the shift of an emitting point in the disk, and found that, in the weak-field

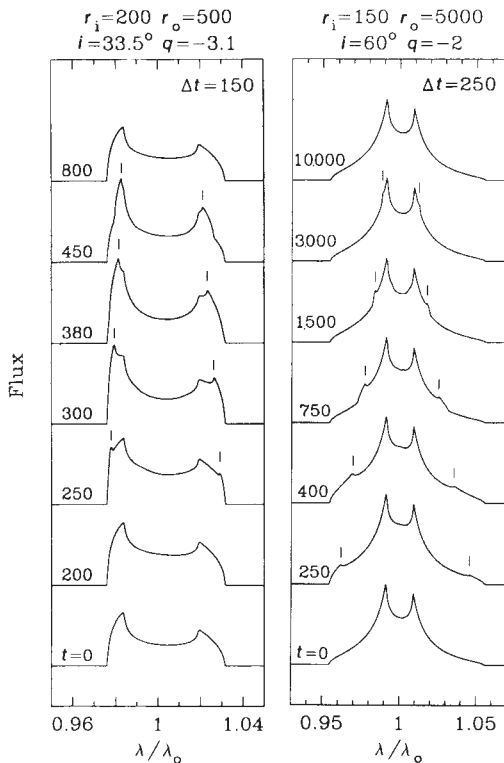


FIG. 2 Response of the line profile from a relativistic accretion disk to flux variations in the central source, as a function of wavelength. The disk parameters in the right and left panels are those used to model the H α profile of Arp 102B (ref. 2) and the H β profile of 3C 390.3 (ref. 3), respectively. The observation spacing required to follow the drift of the red and blue features is $\sim 0.1 - 1.0 (M/10^6 M_\odot)$ days. The variability observed in ref. 16 in the blue wing of the H α profile of Arp 102B with observations separated by ~ 1 yr might be due to the effects discussed here; better coverage is required to confirm this interpretation.

approximation ($r \gg r_s$), the largest blueshift and redshifts in the echo front are produced for

$$\sin \phi_f = \{(\sin i)^{-1} - [(\sin i)^{-2} + 3]^{1/2}\} / 3 \quad (2)$$

that is, angles $\leq 20^\circ$ beyond the line of nodes. The corresponding radius from equation (1) is

$$r_f = c(t - t_0)(1 - \sin \phi_f \sin i)^{-1} - r_s \quad (3)$$

where the last term on the right-hand side was introduced to account approximately for general relativistic corrections at small radii¹⁴. By inserting equations (2) and (3) in

$$(1+z)_f = \left(1 - \frac{3r_s}{2r_f}\right)^{-1/2} \left\{1 \pm \frac{\cos \beta_f}{[2r_f(1 + \tan^2 \xi_f)/r_s - 2]^{1/2}}\right\} \quad (4)$$

where $\cos \beta_f = \cos \phi_f \sin i (1 - \sin^2 \phi_f \sin^2 i)^{-1/2}$ and $\tan \xi_f = \tan \phi_f \cos \beta_f$ (for details see the appendix of ref. 12), the shifts $(1+z)_f$ of the drifting features are good to an accuracy of a few percent over a wide range of disk parameters (see the vertical lines in the figures).

Conversely, by using equations (2) and (4) to fit the observed position of the blue and red drifting features at any given time, it is possible to determine i and r_f . Alternatively, by measuring the stationary line profile, the accretion disk inclination i can be determined together with the parameters r_i , r_o and q (refs 2, 12). These measurements of radii involve scale lengths in units of the Schwarzschild radius (compare equation (4)), implying nothing about the mass of the central object. However, by monitoring the position of the blue or red drifting features at two different times t and t' , the difference of the corresponding radii $r'_f - r_f$ is evaluated also in absolute units through equation (3). The combination of the relative and absolute measurements of $r'_f - r_f$ provides the mass of the central object

$$M = \frac{c^3}{2G} (t - t') \left[\left(\frac{r'_f}{r_s} - \frac{r_f}{r_s} \right) (1 - \sin \phi_f \sin i) \right]^{-1} \quad (5)$$

In the presence of an estimate of the disk inclination i it is enough to monitor the drift of only one of the two features. Another possibility is to obtain an absolute measurement of r_f through the delay between the flare onset time t_0 and the appearance of the blue or red feature at a time t .

Observation of these phenomena in the optical lines of disk-accreting AGNs is within reach of current instrumentation (Fig. 2). Higher-resolution observations ($E/\Delta E \geq 50$) to be carried out with the next generation of X-ray satellites should resolve the characteristic double-horned Fe K-line profiles expected from accreting X-ray sources and allow a probing of the inner disk regions. The number of X-ray photons at the Earth per Schwarzschild radius light-crossing time is a factor of $\geq 10^4$ higher for bright AGNs than for X-ray binaries. Prospects for observing the drifting features from individual source flares on the wings of X-ray lines are therefore far more promising in AGNs. Observations of $\sim 0.1(M/10^7 M_\odot)$ days with the X-ray CCDs to be flown on Astro-D, Jet-X, AXAF and XMM should be able to monitor the Fe K-line profile changes in a number of bright Seyfert galaxies and quasars. Ultimately the observability of individual drifting features will depend on whether sufficiently large flux variations with rise time $\leq r_f/c$ occur in the central source. In the absence of such flares, the continuous aperiodic variability observed in a large number of accreting collapsed objects should allow measurement of the drift of the red and blue features by evaluating cross-correlation delays between variations at different wavelengths in the line wings.

Variable line profiles from accretion disks should provide far more reliable measurements of black-hole masses in AGNs than those available so far.¹⁵ $\square$

2. Chen, K., Halpern, J. P. & Filippenko, A. V. *Astrophys. J.* **339**, 742–751 (1988).
3. Chen, K. & Halpern, J. P. *Astrophys. J.* **344**, 115–124 (1989).
4. Nandra, K., Pounds, K. A., Fabian, A. C. & Rees, M. J. *Mon. Not. R. astr. Soc.* **236**, 39P–46P (1989).
5. Pounds, K. A., Nandra, K., Stewart, G. C. & Leighly, K. *Mon. Not. R. astr. Soc.* **240**, 769–783 (1989).
6. George, I. M., Nandra, K. & Fabian, A. C. *Mon. Not. R. astr. Soc.* **242**, 28P–31P (1990).
7. Barr, P., White, N. E. & Page, C. G. *Mon. Not. R. astr. Soc.* **216**, 65P–70P (1985).
8. van der Woerd, H., White, N. E. & Kahn, S. M. *Astrophys. J.* **344**, 320–324 (1989).
9. White, N. E. *et al.* *Mon. Not. R. astr. Soc.* **218**, 129–138 (1986).
10. Hirano, T. *et al.* *Publ. astr. Soc. Japan* **39**, 619–644 (1987).
11. Kallman, T. & White, N. E. *Astrophys. J.* **341**, 955–960 (1989).
12. Fabian, A. C., Rees, M. J., Stella, L. & White, N. E. *Mon. Not. R. astr. Soc.* **238**, 729–736 (1989).
13. Blandford, R. D. & McKee, C. F. *Astrophys. J.* **255**, 419–439 (1982).
14. Paczynski, B. unpublished (1978) cited in Abramovitz, M. A., Calvani, M. & Nobili, L. *Astrophys. J.* **242**, 722–788 (1980).
15. Wandel, A. & Mushotzky, R. F. *Astrophys. J.* **306**, L61–L65 (1986).
16. Halpern, J. P. & Filippenko, A. V. *Nature* **331**, 46–48 (1988).

ACKNOWLEDGEMENTS. I am grateful to P. Barr, A. Pollock and N. White for useful comments on an earlier version of this letter.

Germanium and silicon in rivers of the Orinoco drainage basin

Richard J. Murnane* & Robert F. Stallard†

* Department of Geological and Geophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA

† US Geological Survey—Water Resources Division, Box 25046, MS-413, Denver Federal Center, Denver, Colorado 80225-0046, USA

CONTINENTAL weathering and mid-ocean-ridge hydrothermal circulation are the two main processes that release dissolved silicon to the ocean. It has been hypothesized¹ that temporal variations in the relative strengths of these two processes might be estimated from the Ge/Si ratio of opal in marine sediments, because the chemistry of germanium is similar to that of silicon^{2–5}. We analysed river water from the Orinoco basin and soils from Venezuela and Panama for germanium and silicon, to determine the relationship between river Ge/Si ratios, drainage basin lithology and chemical-weathering intensity. Our results, presented here, indicate that chemical-weathering intensity, defined as the fraction of original bedrock silicon dissolved during continental weathering, is related to the Ge/Si ratio in river water and is inversely related to river silicon concentrations. Changes in global climate, sea level or patterns of plate-tectonic interactions could change chemical-weathering intensity and complicate the interpretation of the marine opal Ge/Si record.

The chemical properties² and inorganic geochemical pathways^{3–6} of Ge and Si are very similar. Ge substitutes for Si in mineral lattices and tends to have particularly high concentrations in sulphide ore deposits, iron oxide deposits, and coals⁶. In rivers, Ge and Si are present largely in inorganic forms^{3,4}: $\text{Ge}(\text{OH})_4$ and $\text{Si}(\text{OH})_4$, hereafter referred to as Ge_D and Si_D , respectively. The average $\text{Ge}_\text{D}/\text{Si}_\text{D}$ ratio in river water from regions that are not influenced by coal burning is about $0.6 \times 10^{-6} \pm 10\%$ ^{3,4}. This ratio is less than typical bedrock ratios of $\sim 1.5\text{--}2.5 \times 10^{-6}$ (ref. 4).

During chemical weathering of silicate minerals, a portion of Si is retained in durable primary minerals, mostly quartz, and new secondary minerals, mostly clays. We define chemical-weathering intensity, W , as the fraction of total Si dissolved from original bedrock during weathering. This definition refers only to the breakdown of silicate minerals; carbonate and evaporite minerals, which weather congruently, are disregarded. W is similar to the dissolved transport index (DTI) for Si defined by Martin and Meybeck⁷, except that W refers to the weathering process and DTI refers to material being transported by rivers. Under conditions of steady-state erosion, the two indices are equivalent.

Rivers in the Orinoco basin drain geologically diverse regions⁸. The southeastern one-third of the basin is dominated by the Precambrian Guayana Shield. In the western and north-western parts of the basin are the fold-and-thrust terrains of the Andes and the Caribbean marginal systems. These mountains

Received 20 November 1989; accepted 22 February 1990.

1. Perez, E., Penston, M. V., Tadhunter, C., Mediavilla, E. & Moles, M. *Mon. Not. R. astr. Soc.* **230**, 353–362 (1988).

TABLE 1 Soil-profile analyses

Sample	Ge ($\mu\text{mol kg}^{-1}$)	Si (mol kg^{-1})	Ge/Si $\times 10^6$	Σz^{++}
ORS-1 (bedrock) [†]	18.1 $\pm$ 0.1	12.2 $\pm$ 0.3	1.5 $\pm$ 0.1	4.20
ORS-8 (surface soil)	17.4 $\pm$ 0.7	4.6 $\pm$ 0.2	3.8 $\pm$ 0.2	0.016
BCI-10a (bedrock) [‡]	17.2 $\pm$ 1.5	8.4 $\pm$ 0.1	2.0 $\pm$ 0.2	5.82
BCI-1 (surface soil)	25.3 $\pm$ 2.3	5.4 $\pm$ 0.2	4.7 $\pm$ 0.5	0.079
BCI-2 (surface soil)	28.8 $\pm$ 1.0	5.5 $\pm$ 0.2	5.2 $\pm$ 0.3	0.055

Soil samples were fused with lithium metaborate and then dissolved and diluted using 3M HNO₃. After dilution of the concentrated soil solutions, major elements were analysed¹². For Ge analysis, samples were dissolved using HF, H₂SO₄ and HNO₃. Ge was then extracted from HCl solutions using CCl₄, back-extracted using water, and then analysed^{12,18}.

* Total for Na⁺, K⁺, Ca²⁺ and Mg²⁺ in microequivalents per kilogram.

[†] ORS, samples of the Parguaza Granite from 20 km south of Puerto Ayacucho, Amazonas Territory, Venezuela; sample numbers increase towards surface; see Stallard⁸.

[‡] BCI, samples of basalt from the volcanic facies of the Caimito Formation, end of the Barbour Trail, Barro Colorado Island, Panama; sample numbers decrease towards surface.

are being simultaneously uplifted and rapidly eroded. Between these provinces is the foreland basin that receives sediment from the mountain belts. This is a region of alluvial plains of extremely low relief (rarely reaching 200 m) called the Llanos. Rivers in the Llanos can be divided into those that have headwaters in the mountain belts and foothill regions, and those that only drain the Llanos. The sample suite from the Orinoco basin avoids possible problems of anthropogenically elevated Ge_D concentrations associated with coal combustion^{3,4}. We assume that these data indicate natural variability in Ge_D concentrations and in Ge_D/Si_D ratios.

The nature of erosion processes, which we describe in terms of erosion regimes, varies throughout the Orinoco basin. An erosion regime classifies hillslope erosion processes in terms of a continuum between weathering-limited and transport-limited erosion⁸. Erosion is weathering-limited when the capacity of the transport processes exceeds the rate at which material is generated by weathering, whereas erosion is transport-limited when the supply rate of material due to weathering exceeds the removal capacity of transport processes. In the mountain belts, erosion is weathering-limited, and rivers have large dissolved

and suspended loads⁸. The steep relief in the mountains facilitates physical removal of loose soil and rock and constantly exposes fresh material for chemical leaching. In steep areas of the shield, erosion is also weathering-limited, but because bedrock is resistant to chemical weathering, denudation is slow. Erosion is transport-limited in the extremely flat regions on the lowland Guayana Shield and in parts of the Llanos that have deeply weathered alluvial soils. In these regions, chemical leaching of soils and rocks is the predominant weathering mechanism and physical erosion is minor. Rivers in these areas have low Si_D and total cation concentrations, and small suspended loads.

Within tropical river basins chemical-weathering intensity and the chemical compositions of river water and soil are correlated with the erosion regime⁸⁻¹¹. The most intense chemical weathering, which involves almost complete dissolution of Si in bedrock, occurs in very flat regions where denudation is slow (<20 m per Myr) and erosion is transport-limited^{8,9}. Si_D is the predominant dissolved inorganic component even though its concentration is low (<40 $\mu\text{mol l}^{-1}$); Al, Fe-sesquioxides, kaolinite and quartz are thermodynamically unstable¹¹, and soils often are podzolic. Major soluble cation (Na⁺, K⁺, Mg²⁺, Ca²⁺) concentrations in river water are in almost-bedrock proportions, and high concentrations of dissolved Al and Fe (>1 $\mu\text{mol l}^{-1}$) are common¹¹.

Chemical weathering is less intense and denudation is more rapid (10–80 m per Myr) where weathering-limited erosion of elevated shield or transport-limited erosion of alluvial plains is occurring^{8,9}. Although Si_D is no longer a predominant component, Si_D concentrations are greater (40–300 $\mu\text{mol l}^{-1}$). The greatest Si_D concentrations are in rivers that drain mafic rocks. Concentrations of dissolved Al and Fe are low, and major soluble cations are in bedrock proportions. Soils contain Al, Fe-sesquioxides, kaolinite and quartz.

Chemical weathering is least intense and denudation is most rapid (50–5,000 m per Myr) in mountainous areas undergoing weathering-limited erosion. Denudation is most rapid where unstable lithologies such as shales, carbonates and evaporites are eroding^{8,9}. Rivers which have high Si_D concentrations drain the mountains and parts of the Llanos that receive fresh sediment from the mountains. Si_D concentrations vary between 100–500 $\mu\text{mol l}^{-1}$, and soil suites of secondary minerals correlate with the total soluble cation concentration, Σz^{+} (after correction for cyclic salt inputs and halite weathering), of the river water⁸⁻¹¹. In regions where rivers have a Σz^{+} of $\leq 500 \mu\text{eq l}^{-1}$, kaolinite and Fe-sesquioxides are the most important secondary minerals; where Σz^{+} is greater, 2:1 clays (clays with an octahedral layer between two layers of (SiO₄) tetrahedra) are important. The formation of 2:1 clays, which have large Si/Al ratios, seems to limit Si_D¹¹. The inverse correlation between chemical-weathering intensity and the concentrations of Si_D and major soluble cations indicates a first-order control of weathering rates by hillslope erosion processes; the most intense chemical weathering occurs in the flattest regions^{8,9}.

The depletion of Ge_D relative to Si_D, compared with their proportions in bedrock, can be described using a simple partitioning model that relates the release of Ge_D and Si_D to chemical-weathering intensity and bedrock composition. Assuming that there is no storage of Ge_D and Si_D in ground water or lakes and that chemical weathering is sufficiently intense to dissolve all Si and Ge from bedrock, the ratio of Ge_D/Si_D in rivers, R_D , must equal the Ge/Si ratio of bedrock, R_B . In most uncontaminated rivers, R_D , is typically less than R_B . This indicates that Ge is preferentially partitioned in to the solid weathering products, and the Ge/Si ratios in solid weathering products, R_S , must be greater than R_B . Thus, as the intensity of chemical weathering increases, R_D should increase until it equals R_B . Two distinct processes can cause Ge enrichment in solid weathering products. One is the dissolution of phases with a low Ge/Si ratio; the other is the selective retention of Ge during the formation of the secondary phases. When the quantity of

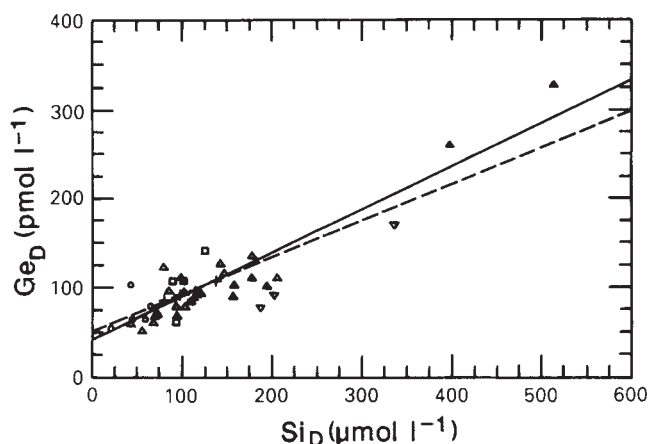


FIG. 1 Ge_D (pmol l^{-1}) compared to Si_D ($\mu\text{mol l}^{-1}$) for all rivers. The solid line represents a least-squares line, $\text{Ge}_D \times 10^{-6} = 39 \pm 3 + 0.49 \pm 0.07 \text{Si}_D$ ($r = 0.89$, 95% confidence). The dashed line is the same as the dashed line in Fig. 2, transformed to the coordinate axes of this figure. Symbols: +, Lower Orinoco mainstem; O, low, flat shield that has felsic bedrock; Δ , elevated shield that has mafic bedrock; $\square$, low, alluvial plains that have deeply weathered soils; $\blacktriangle$, rivers that have headwaters in mountain belts. River samples from the Orinoco basin were collected during October 1984, filtered through 0.45 μm cellulose acetate filters, and stored in the dark before analysis^{12,17}. Most Ge_D analyses used 100-ml sample volumes. Analytical precision is estimated to be $\pm 10\%$.

Si released to solution is small compared to that eroded in the solid form, R_S must equal R_B .

The total mass of Si weathered from original bedrock in a time interval, $Q_{Si:T}$, equals the sum of Si removed in solution, $Q_{Si:D}$, and that removed in a solid form, $Q_{Si:S}$; the same mass-balance relation applied to Ge:

$$Q_{Si:T} = Q_{Si:S} + Q_{Si:D} \quad (1a)$$

$$Q_{Ge:T} = Q_{Ge:S} + Q_{Ge:D} \quad (1b)$$

$R_D = Q_{Ge:D}/Q_{Si:D}$; $R_S = Q_{Ge:S}/Q_{Si:S}$; and $R_B = Q_{Ge:T}/Q_{Si:T}$. From equations (1a, b), R_B also $= (Q_{Ge:S} + Q_{Ge:D})/(Q_{Si:S} + Q_{Si:D})$. Therefore,

$$\begin{aligned} R_B \cdot Q_{Si:T} &= R_B(Q_{Si:S} + Q_{Si:D}) \\ &= R_S \cdot Q_{Si:S} + R_D \cdot Q_{Si:D} \end{aligned} \quad (2)$$

The chemical-weathering intensity, W , is $Q_{Si:D}/Q_{Si:T}$. Equation (2) becomes:

$$R_B = R_S(1 - W) + R_D \cdot W \quad (3)$$

Because of the chemical similarity of Ge and Si, we assume the simplest possible relation between R_S and W , that of a straight line with an intercept of R_B at $W = 0$:

$$R_S = R_B + A \cdot W \quad (4)$$

The slope of the line is an empirical partitioning factor that can be calculated from the data. This linear model does not include the complexities associated with rapid physical erosion, without substantial chemical weathering, of poorly lithified siliceous sediments. Substituting equation (4) into equation (3) we derive an equation for R_D in terms of R_B :

$$R_D = R_B + A(W - 1) \quad (5)$$

Equations (4) and (5) include the essential parameters needed to estimate the chemical-weathering intensity of a river basin when the Ge/Si ratios in bedrock and river water are known. A must be calculated empirically. If A is not zero, R_D depends directly on R_B and on W . When $W = 0$, $R_S = R_B$, and when $W = 1$, $R_D = R_B$.

For rivers within the Orinoco basin, Ge_D concentrations range from 51 pmol l^{-2} to 333 pmol l^{-1} and Si_D concentrations range from $21 \text{ } \mu\text{mol l}^{-1}$ to $510 \text{ } \mu\text{mol l}^{-1}$ (Fig. 1). R_D values range from 0.40×10^{-6} to 2.62×10^{-6} (Fig. 2) and are linearly related to $1/Si_D$. Ge_D and Si_D concentrations and R_D are generally within the range of remote clean rivers^{3,4}, but some rivers have a substantially greater R_D . The slope of the regression line that relates river Ge_D and Si_D concentrations is 0.49×10^{-6} (Fig. 1),

a value slightly less than the average R_D (0.6×10^{-6}) of clean rivers⁴. R_D values also decrease as Σz^{+} increases¹².

Rivers that drain mafic rocks have a distinctive signature and plot at the high- Si_D , low- R_D end of the trends shown in Figs 1 and 2, presumably reflecting the great susceptibility to chemical weathering of mafic minerals. Rivers that drain regions in which erosion is transport-limited have the lowest Si_D concentrations and the greatest R_D , regardless of lithology. Rivers that drain felsic and sedimentary rock plot along the entire trend.

Analyses of soil profiles indicate the nature of changes in solid-phase chemistry caused by weathering. In the soil profile developed on granite (ORS samples in Table 2), the surface sample consists of quartz and cation-deficient secondary minerals (kaolinite, gibbsite and goethite)⁹. The surface sample of the basalt profile (BCI samples in Table 2) contains only secondary minerals (kaolinite, gibbsite, goethite and haematite). Samples obtained from the top of both profiles have Ge/Si ratios that are greater than bedrock ratios. Quartz, the only bedrock mineral remaining in either profile, has relatively small Ge/Si ratios^{6,13}; Ge must therefore be retained preferentially in clays or other weathering products. This retention is consistent with the observation that Ge is concentrated, relative to Si, in minerals, such as clays, that have weakly polymerized silicate tetrahedra⁶. Fe-sesquioxides, which also retain Ge and occur in these soils, may also contribute to the enrichment of Ge.

A range of values for A can be estimated from the river and soil data. If the composition of the surface soil samples is assumed to correspond to that of solids being eroded during intense chemical weathering, $W \approx 1$, then from equation (4), A is calculated to be 2.3×10^{-6} for the granite soil (ORS samples) and 2.7×10^{-6} for the basalt soil (BCI samples). For the river samples, the intercept (0.42×10^{-6}) of the line that relates R_D to $1/Si_D$ in Fig. 2 is assumed to correspond to R_D when $W \approx 0$. Similarly, the largest values of R_D in the river samples are assumed to correspond to the situation where $W \approx 1$, and $R_B \approx R_D$. R_D may be as large as 2.6×10^{-6} ; however, the average value of R_D for rivers that drain the flat shield is 1.7×10^{-6} . If these values are used as a range for R_B at $W \approx 0$, then, from equation (5), a range for A is estimated to be 1.3×10^{-6} to 2.2×10^{-6} .

Equation (5) can be solved for W :

$$W = 1 - (R_B - R_D)/A \quad (6)$$

The value of W calculated for the most downstream Orinoco sample at Punta Cuchillo near Ciudad Bolívar, Venezuela ($R_D = 0.90 \times 10^{-6}$) ranges from 0.23 ($R_B = 2.6 \times 10^{-6}$, $A = 2.2 \times 10^{-6}$) to 0.38 ($R_B = 1.7 \times 10^{-6}$, $A = 1.3 \times 10^{-6}$). Using the same A and R_B and the estimated value of 0.6×10^{-6} for the average R_D of clean rivers⁴, a range of W for world-average rivers is 0.09 to 0.15. This range is substantially greater than the value of 0.02 for W estimated by Martin and Maybeck⁷ for world rivers. The difference is not unreasonable, because much of riverine solid load probably is derived from the physical erosion of poorly lithified sediment, and is therefore not accounted for by this model of chemical erosion. Using $W = 0.02$, predicted world-average river R_D is 0.4×10^{-6} to 0.5×10^{-6} . This R_D is close to the smallest R_D of our sampled rivers, $\sim 0.5 \times 10^{-6}$.

Within the Orinoco basin, chemical-weathering intensity within river catchments is related inversely to Si_D concentration in the rivers. This is shown by substituting the linear relation in Fig. 2 into equation (6) using values for R_B and A that were estimated previously by substitution of the same linear relation into equation (5):

$$\begin{aligned} W &= (\text{slope of relation})/(A \cdot Si_D) \\ &= 20/Si_D \text{ to } 40/Si_D \end{aligned} \quad (7)$$

Meybeck¹⁴ observed that Si_D in rivers from cooler climatic regions is typically less than Si_D in rivers from warmer regions. He advanced the hypothesis that this is, in part, caused by a decrease in the intensity of chemical weathering in cooler areas.

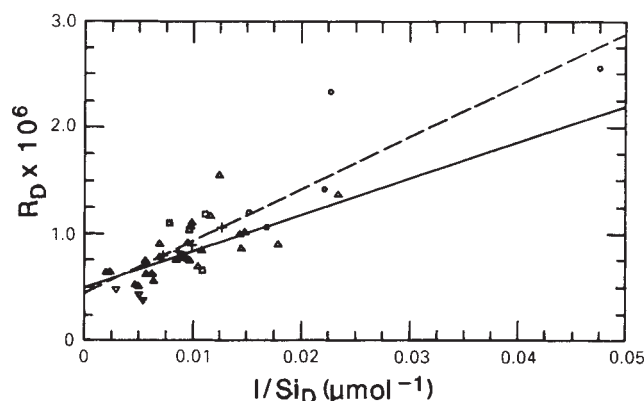


FIG. 2 River $R_D \times 10^6$ ($Ge_D/Si_D \times 10^6$) compared to inverse Si_D concentration. The dashed line represents a least-squares line, $R_D \times 10^6 = 0.42 \pm 0.03 + 40 \pm 9/Si_D$ ($r = 0.87$, 95% confidence). The solid line is the same as the solid line in Fig. 1, transformed to the coordinate axes of this figure. Symbols are the same as in Fig. 1. Linear relations of the type seen in Fig. 1 are preserved in the transformation used for the data in this figure.

Equation (7) is inconsistent with this hypothesis, indicating that the slope of equation (7) could be climate dependent. Consequently, it would be interesting to test whether equivalent relations are valid for other large, climatologically homogeneous, uncontaminated river basins.

These data indicate that the Si_D flux and Ge_D/Si_D ratio of fluvial inputs to the ocean could be affected by changes in either chemical-weathering intensity or in the relative extents of different lithologies exposed to weathering over the Earth. A greater contribution of Si_D and Ge_D to the oceans from regions that are undergoing transport-limited erosion relative to those undergoing weathering-limited erosion could cause an increase in the Ge_D/Si_D of rivers. Situations involving changes of climate or plate-tectonic patterns which could accomplish this include: increasing the exposure of shields relative to mountain belts, and the warming of global climate causing more intense chemical weathering. In contrast, more extensive exposure of mafic rocks would decrease ratios. Because of these ambiguities, it would be difficult to distinguish, without additional information, variations in oceanic Ge_D/Si_D ratios caused by changes in average river Ge_D/Si_D from those caused by changes in other aspects of the geochemical cycles of Ge and Si.^{3-5,12,15,16} □

Received 17 October 1989; accepted 15 February 1990.

1. Froelich, P. N. & Andreae, M. O. *Science* **213**, 205–207 (1981).
2. Cotton, F. A. & Wilkinson, G. *Advanced Inorganic Chemistry* (Wiley-Interscience, New York, 1972).
3. Froelich, P. N., Hambrick, G. A., Andreae, M. O., Mortlock, R. A. & Edmond, J. *J. geophys. Res.* **90**, 1133–1141 (1985).
4. Mortlock, R. A. & Froelich, P. N. *Geochim. cosmochim. Acta* **51**, 2075–2082 (1987).
5. Murnane, R. J., Leslie, B., Hammond, D. A., Stallard, R. F. *Geochim. cosmochim. Acta* **53**, 2873–2882 (1989).
6. Bernstein, L. R. *Geochim. cosmochim. Acta* **49**, 2409–2422 (1985).
7. Martin, J. M. & Meybeck, M. *Marine Chem.* **7**, 173–206 (1979).
8. Stallard, R. F. in *The Chemistry of Weathering* (ed. Drever, J. I.) 293–316 (Reidel, Dordrecht, 1985).
9. Stallard, R. F. in *Physical and Chemical Weathering in Geochemical Cycles* (eds Lerman, A. & Meybeck, M.) 225–246 (Kluwer Academic, Dordrecht, 1988).
10. Stallard, R. F. & Edmond, J. M. *J. geophys. Res.* **88**, 9671–9688 (1983).
11. Stallard, R. F. & Edmond, J. M. *J. geophys. Res.* **92**, 8293–8302 (1987).
12. Murnane, R. J. thesis, Princeton Univ. (1988).
13. Novokhatskiy, I. P., Kalinin, S. K. & Zamayina, G. M. *Geochim. Int.* **4**, 1192–1196 (1967).
14. Meybeck, M. *Rev. Geol. Dyn. Geogr. Phys.* **21**, 215–246 (1979).
15. Mortlock, R. A. & Froelich, P. N. *Science* **231**, 43–45 (1986).
16. Murnane, R. J. & Stallard, R. F. *Paleoceanography* **3**, 461–469 (1987).
17. Andreae, M. O. & Froelich, P. N. *Analyt. Chem.* **553**, 287–291 (1981).
18. Schneider, W. A. Jr & Sandell, E. B. *Mikrochim. Acta* **2**, 263–268 (1954).

ACKNOWLEDGMENTS. Funding for work on the Orinoco River came from the NSF and the Venezuelan Ministerio del Ambiente y de los Recursos Naturales Renovables. Proyecto Orinoco-Apure. We thank Dr G. C. Finol, A. Mejia, C. M. Delgado Z., J. Sainz and D. P. Hernandez, for their support of the field program. R. H. Meade, C. F. Nordin, and C. C. Cranston (USGS); J. M. Edmond and B. Grant (MIT); O. Alvarado R., L. H. Bravo, and O. Hernandez (MARNR); C. E. Yanez C. (UC); and J. Paulini and R. Herrera (IVIC) participated in some of the field work.

Fossil evidence for the location of the Precambrian/Cambrian boundary in Morocco

Andrew Latham & Robert Riding

Department of Geology, University of Wales College of Cardiff, Cardiff CF1 3YE, UK

THE Precambrian/Cambrian boundary is one of the most important points of the geological timescale, yet it remains stratigraphically undefined¹. Most of the candidate sections are in carbonate rocks², so their carbon isotope stratigraphy can be used to infer the timing of changes in ocean chemistry³. These $\delta^{13}\text{C}$ profiles have been used to correlate Precambrian/Cambrian boundary sections where palaeontological information is scarce^{4–7}. In the Anti-Atlas mountains of Morocco, two distinct positive $\delta^{13}\text{C}$ excursions, about 1,200 m apart in the sequence, have been detected⁸. On the basis of minimal palaeontological information the lower of these two excursions was correlated with the base of the Vendian (latest Precambrian), and the upper with the Tommotian/Atdabanian boundary⁸. We present here new palaeontological data, from calcified cyanobacteria⁹ and trace fossils, indicating that the consensus view on the position of the Precambrian/Cambrian boundary in this sequence is about 1,000 m too high (Fig. 1).

A thick, apparently complete sequence of platform carbonates spanning the Precambrian/Cambrian boundary is well exposed over long distances in the western Anti-Atlas Mountains of Morocco¹⁰. The best known section is at Tiout, 80 km east of Agadir^{11,12}. Archaeocyaths^{13,14} and trilobites¹⁵ indicate a low to mid-Atdabanian age for the upper 120 m of the Calcaire Supérieur (defined as the 'Transition Layers'), whereas the underlying Série Lie de Vin has, on the basis of lithology¹⁶ and stromatolites and thrombolites^{17,18}, been dated as of Vendian age. Consequently, the consensus of work on this sequence has placed the Precambrian/Cambrian boundary near the top of the Série Lie de Vin, or within the base of the Calcaire Supérieur^{1,8,12,17–18}. An alternative, minority view—which is supported by our new data—is that the Série Lie de Vin stromatolites and thrombolites are in fact Cambrian in age²⁰, and that the Precambrian/Cambrian boundary is near the top of the underlying Dolomie Inférieure. Lower in the sequence, a Precambrian age is indicated by an Ediacaran-type soft-bodied fauna²¹ and Upper

Riphean acritarchs¹⁸ in the Série de Base. Tilloids occur in the Tiddiline Formation of the underlying Precambrian basement²². Volcanic rocks of the Ouazazate Series, which are unconformably overlain by the Série de Base over 100 km east of Tiout, have yielded U–Pb ages of 563 ± 20 Myr (ref. 23).

The new fossil discoveries reported here include the probable cyanobacteria *Tarthinia* Drosdova and *Renalcis* Vologdin (Fig. 2a, b), which construct the clotted microfabric of thrombolites throughout the Série Lie de Vin at Tiout⁹ (Fig. 1). The association of *Tarthinia* and *Renalcis* has not been found in strata older than Tommotian²⁴. A single occurrence of the probable cyanobacterium *Kordephyton* Radugin and Stepanova (Fig. 2c) in a thrombolite 750 m above the base of the Série Lie de Vin at Tiout is consistent with a late Tommotian or younger age for the upper 200 m of the Lie de Vin, because *Kordephyton* does not occur in strata below the *Dokidocyathus lenaicus* zone of the late Tommotian in Siberia and Mongolia²⁵. Trace fossils, not previously found in the Série Lie de Vin, include *Arenicolites* Salter, *Diplocraterion* Torrel and *Planolites* Nicholens, and occur at several horizons between 650–750 m above the base of the series at Tiout. *Diplocraterion* (Fig. 2d) is the most biostratigraphically useful form because it is not known to occur in earliest Cambrian rocks²⁶, and it therefore probably signifies a late Tommotian or younger age for the upper part of the Série Lie de Vin. These results thus confirm that the Série Lie de Vin is of Cambrian age.

According to the conventional view of the position of the Precambrian/Cambrian boundary in the sequence, the lower positive $\delta^{13}\text{C}$ excursion, near the top of the Dolomie Inférieure, is correlated with the base of the Vendian⁸. But we now see that the Cambrian fossils (*Tarthinia*) lowest in the sequence are found from 50 m above the base of the Série Lie de Vin at Tiout; that is, only 150 m above this excursion. Whole-rock isotope analyses of carbonates may contain diagenetic influences²⁷, and primary signals can reflect local variation²⁸. Thus, there is still no confirmation that global changes in ocean composition are evident in published isotope curves across the Precambrian/Cambrian boundary. Accepting these reservations, if we take the available information at face value and assume it to reflect global changes, the lower excursion can be re-correlated with positive $\delta^{13}\text{C}$ excursions within the Yudoma Formation of Siberia⁵ and within the Krol D Member of the Lesser Himalaya⁶, at positions close to or just below the Precambrian/Cambrian boundary. A smaller positive $\delta^{13}\text{C}$ excursion is also known in a carbonate sequence close to the Precambrian/Cambrian boundary in Svalbard, but this section does not contain late Vendian

marine rocks, and the positive isotope ratios are in lacustrine carbonates^{4,29}. An isotope profile of the Yangtse Gorges section in China does not reveal a positive $\delta^{13}\text{C}$ excursion in the uppermost Vendian⁷, but the profile does show a negative shift in $\delta^{13}\text{C}$ values below strata containing Atdabanian trilobites³⁰, which may correlate with negative $\delta^{13}\text{C}$ values in the Moroccan, Siberian and Himalayan sequences. Our new biostratigraphic data, together with this tentative re-calibration of the isotope

curve, thus indicate that the Precambrian/Cambrian boundary should now be placed at the top of, or within the upper 100 m of, the Dolomie Inférieure (Fig. 1).

The upper positive $\delta^{13}\text{C}$ excursion at Tiout, within the Calcaire Supérieur, occurs 300 m above occurrences of *Kordephyton* and *Diplocraterion* (which indicate a maximum age of late Tommotian), and 100 m below mid-Atdabanian archaeocyaths¹⁴. It is therefore within the part of the sequence

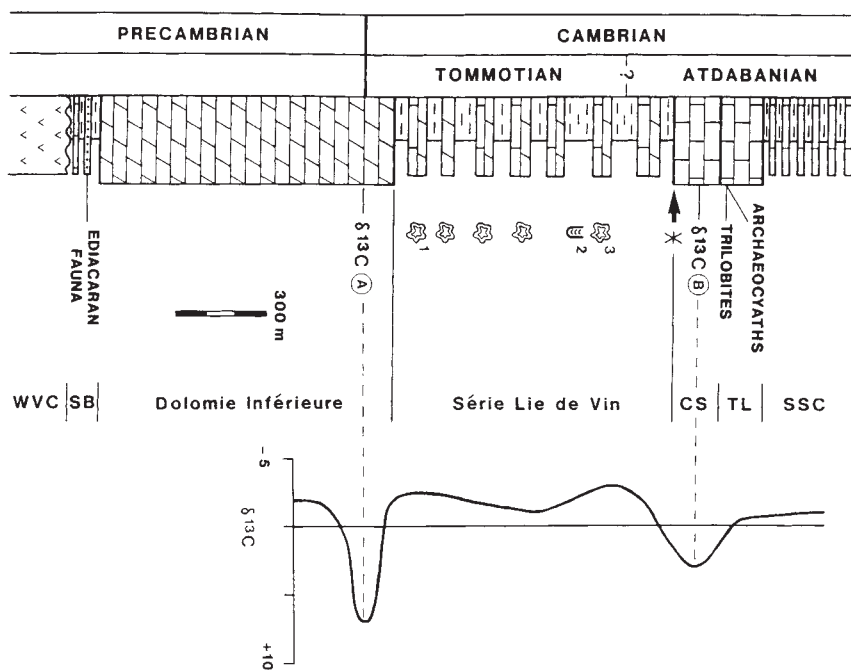
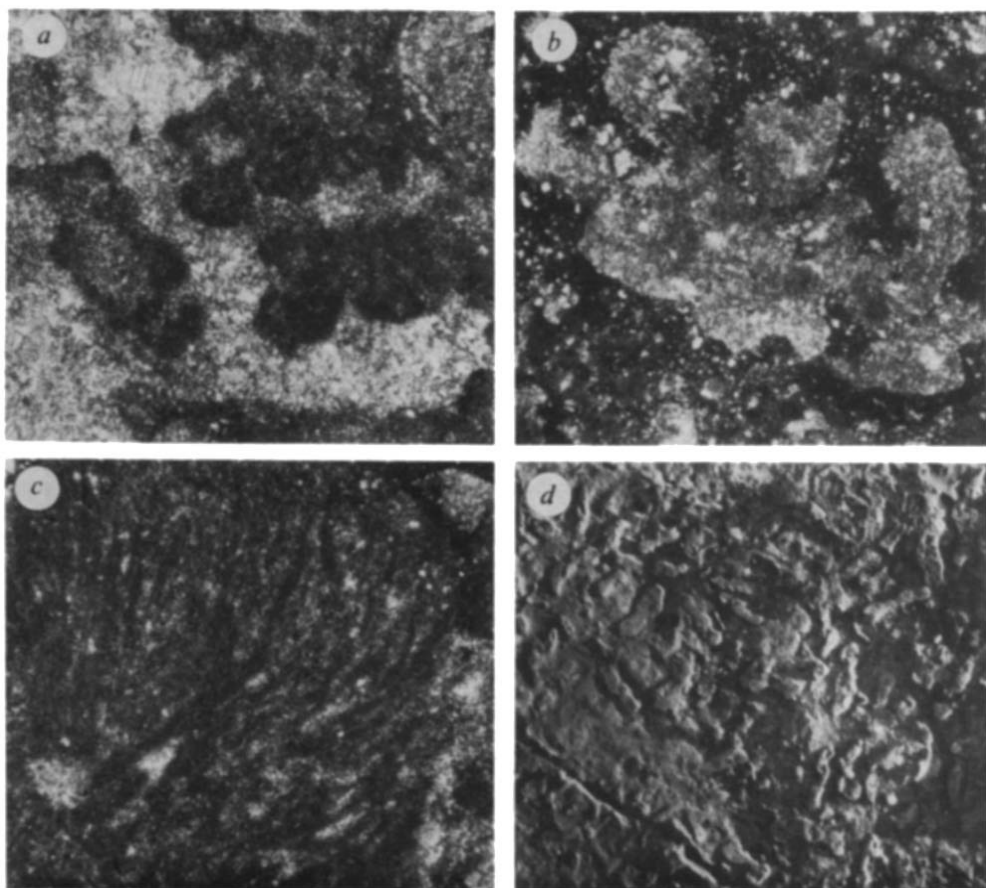


FIG. 1 Stratigraphy of the Tiout section, western Anti-Atlas Mountains, with revised isotope profile. $\delta^{13}\text{C}$ (A) and (B), lower and upper carbon isotope excursions⁸; 1, lowest occurrence of *Tarthinia/Renalcis*-bearing thrombolites; 2, *Diplocraterion*; 3, *Kordephyton/Renalcis/Tarthinia*-bearing thrombolite. WVC, Wawfengha Volcanic Complex; SB, Série de Base; CS, Calcaire Supérieur; TL, Transition Layers; SSC, Série Schisto-Calcaire. On the basis of the fossils reported here, the position of the Precambrian/Cambrian boundary is placed below the Série Lie de Vin, whereas on the basis of the positive isotope excursion it is placed near the top of the Dolomie Inférieure.

FIG. 2 *a*, *Renalcis* sp., Upper Série Lie de Vin, Tiout (Lower Cambrian). Enlargement, $\times 45$. *b*, *Tarthinia* sp., Upper Série Lie de Vin, Tiout (Lower Cambrian); $\times 45$. *c*, *Kordephyton* sp., Upper Série Lie de Vin, Tiout (Lower Cambrian); $\times 45$. *d*, *Diplocraterion* sp., top view; illustrates dumbbell-shaped transverse sections. Upper Série Lie de Vin, Tiout (Lower Cambrian); $\times 0.3$.



that is likely to contain the Tommotian/Atdabanian boundary. More precise correlation of this excursion with the base of the Atdabanian, as suggested by Tucker⁸, cannot be achieved using the fossils that have so far been discovered here as none of them is exclusively Tommotian. A profile through the basal Cambrian of the Australian Flinders Ranges³¹ reveals a positive $\delta^{13}\text{C}$ excursion near the first appearance of archaeocyaths within the Atdabanian, which may correlate with the upper Tiout $\delta^{13}\text{C}$ excursion in the Calcaire Supérieur. Further correlation of this excursion at Tiout is not yet possible, because other $\delta^{13}\text{C}$ profiles through Tommotian and Atdabanian parts of Precambrian/Cambrian boundary sequences have not been reported.

These new microfossil and trace fossil data for the Tiout sequence, together with the re-calibrated isotope curve, identify the position of the Precambrian/Cambrian boundary about 1,000 m lower than was previously generally recognized¹, and they require revision of correlations of $\delta^{13}\text{C}$ profiles, assuming that these reflect global changes. The generally light $\delta^{13}\text{C}$ values throughout the Série Lie de Vin⁸ cannot now be correlated with light values in the Vendian of other sections^{4,5} but correspond instead with light $\delta^{13}\text{C}$ in the Tommotian of Australian³¹. As the quest for a Precambrian/Cambrian boundary stratotype increasingly combines chemical and physical methods of correlation (including magnetostratigraphy and radiometric dating, as well as stable isotopes) with palaeontological information, thick carbonate sequences across the boundary, of which the Dolomie Inférieure of the northwestern Anti-Atlas is a prime example, are likely to assume increasing importance despite their relatively limited potential for biostratigraphy alone. □

Received 17 November 1989; accepted 23 February 1990.

1. Cowie, J. W. & Brasier, M. D. *Oxford Monogr. Geol. Geophys.* **12** (Clarendon, Oxford, 1989).
2. Cowie, J. W. *Episodes* **8**, 93–97 (1985).
3. Veizer, J., Holser, W. T. & Wilgus, C. K. *Geochim. cosmochim. Acta* **44**, 579–587 (1980).
4. Knoll, A. H., Hayes, J. M., Kaufman, A. J., Swett, K. & Lambert, I. B. *Nature* **321**, 832–838 (1986).
5. Magaritz, M., Holser, W. T. & Kirschvink, J. L. *Nature* **320**, 258–259 (1986).
6. Aharon, P., Schildowski, M. & Singh, I. B. *Nature* **327**, 699–702 (1987).
7. Lambert, I. B., Walter, M. R., Wenlong, Z., Songnian, L. & Guogan, M. *Nature* **325**, 140–142 (1987).
8. Tucker, M. E. *Nature* **319**, 48–50 (1986).
9. Latham, A. J. & Riding, R. *Proc. Int. Ass. Sedim. Eur. Meet. Leuven* 127–128 (Int. Ass. Sedim., 1988).
10. Choubert, G. *Congr. Geol. Int. 19e, Alger, Monogr. region* **6**, 77–194 (1952).
11. Schmitt, M. O. & Monninger, W. E. W. in *Fossil Algae* (ed. Flügel, E.) 80–85 (Springer, Berlin, 1977).
12. Monninger, W. E. W. *Arb. Paläont. Inst. Würzburg* **1**, 1–289.
13. Debrenne, F. *Notes & M. Serv. géol. Maroc* **179**, 1–265 (1964).
14. Debrenne, F. & Debrenne, M. *Geol. Mag.* **115**, 101–119 (1978).
15. Szalay, K. *Geol. Mag.* **115**, 83–94 (1978).
16. Hupé, P. *Centre natn. Rech. sci. Paris* **76**, 163 (1958).
17. Schmitt, M. O. *Arb. Paläont. Inst. Würzburg* **2**, 1–188 (1979).
18. Choubert, G., Faure-Muret, A. & Timofeev, B. K. *C. r. hebdo. Séanc. Acad. Sci. Paris* **288D**, 191–194 (1979).
19. Rozanov, A. Yu. *Tr. Akad. Nauk SSSR* **206**, 102–104 (1969).
20. Bertrand-Sarfati, J. *Newslett. Stratigr.* **10**, 20–26 (1981).
21. Houzay, J.-P. *Géol. Méditerranée* **6**, 379–384 (1979).
22. Leblanc, M. in *Earth's Pre-Pleistocene Glacial Record* (eds Hambrey, M. J. & Harland, W. B.) 120–122 (Cambridge University Press, 1981).
23. Mifdal, A. & Peucat, J.-J. *Sci. Géol., Bull.* **38**, 185–200 (1985).
24. Riding, R. & Voronova, L. in *Paleoecology* (eds Toomey, D. F. & Nitecki, M. H.) 56–78 (Springer, Berlin, 1985).
25. Riding, R. & Voronova, L. *Geol. Mag.* **121**, 205–210 (1984).
26. Crimes, T. P. *Geol. Mag.* **124**, 97–119 (1987).
27. Fairchild, I. J., 28th Int. Geol. Congr., Washington 1989. *Abstr.* **1**: 470 (1989).
28. Schildowski, M., Eichmann, R. & Junge, C. E. *Geochim. cosmochim. Acta* **40**, 449–456 (1976).
29. Fairchild, I. J. & Hambrey, M. J. *Precambrian Res.* **26**, 111–168 (1984).
30. Hsu, K. J. *et al. Nature* **316**, 809–811 (1985).
31. Tucker, M. E. *Terra Nova* **1**, 573–582 (1990).

ACKNOWLEDGEMENTS. We thank P. Aharon, M. Brasier, J. Cowie, F. Debrenne and M. Tucker for helpful comments on the manuscript, P. Crimes for trace fossil identifications, and the Ministère de l'Énergie et des Mines, Rabat, for permission to work in Morocco. A.L.'s work was supported by NERC.

Bounds on global dynamic topography from Phanerozoic flooding of continental platforms

Michael Gurnis

Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109-1063, USA

THE movement of continents with respect to a large-scale pattern of dynamic topography and geoid, imposed by convection in the mantle, must contribute to the flooding of continental platforms. Here I investigate this phenomenon, using a one-dimensional model in which a continent moves from a high to a low of dynamic topography (and geoid), and in the process is partially exposed and then flooded. If the dynamic topography is greater than about 150 metres, the model continent is flooded by more than 30%—the maximum amount of flooding experienced by North America during the entire Phanerozoic eon¹. The model suggests that a large-scale pattern of dynamic topography must have an amplitude of less than 150 metres, and that the admittance, the ratio of geoid to dynamic topography, may be greater than 0.3. Recent models of global mantle dynamics^{2,3} which predict the long-wavelength geoid from mantle seismic structure are apparently inconsistent with Phanerozoic flooding.

After the acceptance of plate tectonics, it was noted that the late Cretaceous marine transgression onto the continents was correlated with the 'volume' of oceanic ridges⁴. This led to the hypothesis that when spreading rates and/or length of ridges increase, ocean basins become more shallow and sea level rises⁵. This is unlikely to be the only type of control exerted by the mantle and tectonic plates on long-term eustatic sea-level fluctuations, because density anomalies in the deep mantle (the existence of which is revealed by anomalies in seismic travel times^{3,6,7}, could have caused the continents to bob up and down as they drifted around the surface of the Earth. For a model

that is consistent with the observed geoid, long-wavelength density variations inferred from body-wave seismic tomography give rise to global-scale variations in dynamic topography on the order of one kilometre^{2,3}. Recently, Cazenave *et al.*⁸ were able to extract a global residual topography, of ~300 m amplitude, by correcting for shallow density variations in the lithosphere. These corrections include the removal of contributions due to cooling of the oceanic lithosphere with increasing age⁹ and the subtraction of a mean baseline for stable continental platforms. The residual topography is dominated by a degree-two pattern which is strongly correlated with the geoid⁸.

Another approach to constraining dynamic topography arises from the effect on sea level of the continental 'bobbing' described above: by requiring continental flooding to be less than the observed areal coverage of marine deposits, an upper bound can be placed on the magnitude of dynamic topography. In the one-dimensional model used here, a continent is moved over a surface that has a dynamic topography comprised of a single wavelength. Because of the ambiguity in the data, however, essentially any wavelength will result in periodic flooding; here I am concerned only with wavelengths larger than the scale of continents. In this model, dynamic topography, H , and geoid, N , vary as $H_0 \sin 2\pi x$ and $N_0 \sin 2\pi x$ respectively. Here x is non-dimensional and varies between 0 and 1. The admittance, Z , is defined as N_0/H_0 . I have used three values of N_0 : 0, 50 and 100 m; in comparison, the degree-two and order-two geoid deduced in ref. 2 is ~50 m. The continental crust has a thickness W_c which is symmetrical about its centre at x_c . Two types of crustal thickness variations were used, one leading to a linear hypsometry and the other leading to the average present-day continental hypsometry¹⁰ (solid circles, Fig. 1); the latter is hereafter referred to as the 'world' hypsometry data set. The present range in continental hypsometric curves is indicated in Fig. 1 as the shaded area surrounding the 'world' data set. The total height variation for the linear hypsometric type of continent is h_c . The centre of the continent is moved from the peak of dynamic topography at $x = 0.25$ to the minimum at $x = 0.75$. At each value of x , a constant volume of water is placed on the

system. Starting from the topographically lowest point, water is added to the system such that each mass column is in isostatic equilibrium with respect to the dynamic topography; the water surface follows the shape of the geoid. The system was solved by isostatically balancing columns made up of water, crust and mantle with densities of 1,000, 2,700 and 3,300 kg m⁻³ respectively. For models with linear hypsometries, the volume of water added was such that the continent was flooded by 50% ($f_c = 0.50$) when $N_0 = 0$ and $Z = \infty$; this corresponds to a flat water surface with no dynamic topography. For models with the average observed hypsometry, f_c was taken to be 0.18—the average fraction by which North America was flooded over the Phanerozoic¹—when $N_0 = 0$ and $Z = \infty$. Flooding minima and maxima occur over the lowest and highest points in the geoid, for any Z . The fractional change of continent flooded, Δf_c , is defined as the absolute difference in flooding between $x_c = 0.25$ and $x_c = 0.75$. Results are shown in Fig. 2 for the case with 'world' hypsometry, $Z = 0.5$, and $N_0 = 100$ m for $x_c = 0.25$ (top) and 0.75 (bottom). Although the continent is resting at the geoid high at $x_c = 0.25$, it is more exposed than at $x_c = 0.75$ because the dynamic topography has twice the amplitude of the geoid and thus 'pushes' the continent above the sea surface. The change in flooding (Δf_c) between these two cases is 0.26 and the mean flooding is 0.18.

Flooding is strongly influenced by the admittance and the geoid, as summarized in Fig. 3. For a model Earth with little or no dynamic topography (large Z) and a non-zero geoid, the continent is flooded more extensively at the geoid high than at the geoid low. As the admittance diminishes and the amplitude of dynamic topography increases, progressively less flooding occurs over the geoid high and progressively more over the geoid low. When $Z = 1$, the effects of flooding at the geoid high and exposure at the topographic high cancel as the continent moves from geoid high to low; this corresponds to the minimum in the curve of Z against Δf_c in Fig. 3. As Z falls below unity, the amplitude of dynamic topography increases, Δf_c increases,

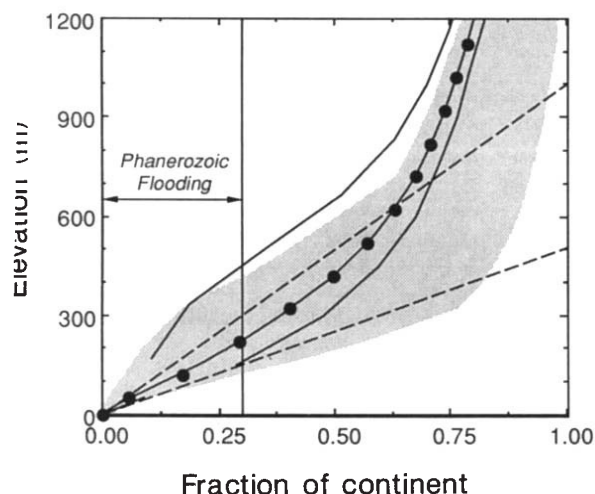


FIG. 1 Observed and calculated hypsometries. The horizontal axis is the fraction of continental area below a given elevation. Solid circles denote the 'world' data set of Harrison *et al.*¹⁰; this is the average of all present-day continents. The extremes of this data set are indicated by the shaded area; the upper edge is defined by the hypsometric curves for Africa and Asia, and the lower by Europe. The area below sea level has been corrected for sea-water loading; the curve has been shifted such that the minimum is at the origin. The data set was fitted with a fifth-order polynomial (solid curve passing through the solid circles) which was then used in the calculations. In the models, the hypsometry is also influenced by dynamic topography, and the extreme variations are shown by the solid lines bounding the 'world' average for a case with $Z = 0.5$ and $N_0 = 100$ m (see text). The dashed curves are two linear hypsometries. The observed range in the fraction of North America flooded over the entire Phanerozoic¹ is also indicated.

and the continent becomes more exposed at the geoid high and more flooded at the geoid lows (compare Fig. 2). If the 'world' hypsometry is used (dotted line *a* in Fig. 3), the range of flooding exceeds 0.30 when Z falls below ~ 0.45 (for $N_0 = 100$ m). North America (the only platform for which a reasonably complete flooding record is available) experienced a maximum flooding range of 30% over the Phanerozoic¹. When the amplitude of dynamic topography is greater than ~ 200 m, flooding in the model is inconsistent with that observed. If N_0 is decreased to 50 m, the amount of flooding also decreases for a given Z (dashed line *b* in Fig. 3); the flooding exceeds 30% when the admittance is less than 0.3. Finally, if there is no geoid ($Z = 0$), the maximum permissible flooding is exceeded when $H_0 > 110$ m. Given that the true geoid should lie in the range 0–100 m, the model strongly suggests that the amplitude of large-scale

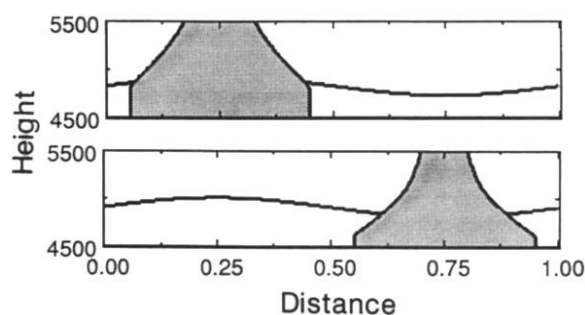


FIG. 2 Total topography and sea surface for the case $Z = 0.5$ and $N_0 = 100$ m at $x_c = 0.25$ (top) and 0.75 (bottom). The topography of the continent (shaded) is the result of isostatic topography (caused by crustal thickness variations), dynamic topography and sea-water loading. The continent becomes more flooded when positioned over a dynamic topography and geoid low. The original hypsometry, without dynamic topography and water loading, would follow the observed hypsometry (the 'world' data set) as shown in Fig. 1. Elevations above 5,500 m have been truncated.

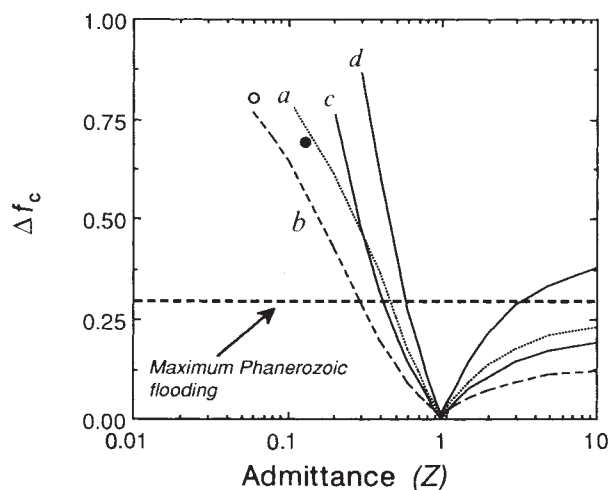


FIG. 3 Total range in fraction of continent flooded, Δf_c , versus admittance, Z , for two cases with the 'world' hypsometry of Harrison *et al.*¹⁰. Curve *a* is for $N_0 = 100$ m and curve *b* is for 50 m. The two solid curves are for the linear hypsometries shown in Fig. 1: *c* with $h_c = 1,000$ m and *d* with $h_c = 500$ m. The 'Maximum Phanerozoic flooding' curve corresponds to the range of flooding that North America experienced during the Phanerozoic¹. In all models, $w_c = 0.4$ (approximately the ratio of continental area (plus their margins) divided by the surface area of the Earth). Two specific spherical dynamic models that predict the observed geoid have been explored: the solid circle denotes model CC-U10 from ref. 2 and the open circle denotes model W4 from ref. 3; both models assume whole-mantle convection with a moderate increase in viscosity from upper to lower mantle.

dynamic topography is $<150 \pm 50$ m. In addition, I have studied two specific cases from the work of Hager *et al.*^{2,3}, where the geoid is calculated using mantle density anomalies resolved through seismic tomography; as discussed above, these models have about 600–1,000 m of dynamic topography and predict flooding (Fig. 3, circles) in excess of 70%—inconsistent with the extent of observed marine deposits.

Flooding is also influenced strongly by hypsometry; this is demonstrated in Fig. 3, where curves *a*, *c* and *d* differ only in respect to their intrinsic hypsometries. Steeper platform hypsometries (larger h_c) result in smaller Δf_c for a given *Z*. Hypsometries evolve^{11,12} as continents are subjected to different tectonic environments at their margins, such that platform slope varies amongst the continents. With the exception of Africa, the variation is at present less than a factor of two¹¹. An uncertainty of this magnitude is represented by the solid lines in Fig. 3: the minimum permissible admittance decreases from 0.55 to 0.40 when platform slope is increased by a factor of two. If the intrinsic variation in average hypsometry over the Phanerozoic has been as great as the present variation between continents¹², the bound on *Z* should vary only by about ± 0.1 if *Z* is in the range 0.3–0.5. But such variations in hypsometry are accounted for implicitly because the total continental hypsometry varies as the continent moves over the undulations in dynamic topography. This variation is shown for a particular model (*Z* = 0.5 and N_0 = 100 m) in Fig. 1 as the solid lines that bound the 'world' hypsometry (solid circles). This variation can be as large as the variation currently observed in the hypsometries between individual continents (shaded area in Fig. 1), and even the slope of the steepest platform (Africa) can be exceeded by the total modelled hypsometry.

The upper bound on a large-scale pattern of dynamic topography is dependent on three fundamental assumptions. (1) The flooding record of North America is representative of all continents. (2) The observed flooding represents the maximum areal coverage of inland seas. (3) A continent experienced complete excursions from a dynamic topography high to a low at least once during the Phanerozoic. Although the flooding-record compilations are not as complete as those for North America, only China (individually) has flooded by significantly more than 30% (ref. 11). If all continents are averaged together, the data summarized by Algeo and Wilkinson¹¹ shows that the average flooding at any given time did not exceed 30%. Because marine deposits are eroded during exposure¹ and rapid transgressive/regressive cycles may not have had time to leave recognizable deposits, the observation of the fraction of platform now covered with marine deposits may not supply an upper bound on flooding. This ambiguity is unavoidable given our understanding of continental stratigraphy. Complete excursion from dynamic topography high to low is a reasonable assumption given that most continents have apparently moved from a geoid high to a geoid low over the last 150 Myr following the breakup of Pangea¹³. In addition, recent theoretical modelling of plate-mantle interaction shows that continental plates consistently make complete excursions from dynamic topography highs to lows^{14,15}.

The models explored here purposely neglect other sources of large-amplitude sea level fluctuations, the most important being variations in oceanic ridge spreading. Incorporation of other mechanisms probably would only tend to increase the range of flooding—the estimated bound on the admittance remains a lower one. The effect on flooding history of restricting the motion of a single continent to one dimension is uncertain, and can be addressed only by moving continental and oceanic plates with their observed palaeo-positions over the surface of a sphere with different admittances and dynamic topographies. Comparison of such models with the observed spatial patterns of marine deposits may reveal the shape of a long-wavelength component of dynamic topography. Such a datum would provide a fundamental constraint on models of mantle convection^{2,3}. □

Received 13 October 1989; accepted 23 February 1990.

1. Wise, D. U. in *The Geology of Continental Margins* (eds Burk, C. A. & Drake, C. L.) 45–58 (Springer, New York, 1974).
2. Hager, B. H., Clayton, R. W., Richards, M. A., Comer, R. P. & Dziewonski, A. M. *Nature* **313**, 541–545 (1985).
3. Hager, B. H. & Clayton, R. W. in *Mantle Convection* (ed. Peltier, R. W.) 657–763 (1989).
4. Hays, J. D. & Pitman, W. C. *Nature* **246**, 18–22 (1973).
5. Kominz, M. A. *Am. Ass. Petrol. Geol. Mem.* **36**, 109–127 (1984).
6. Dziewonski, A. M. *J. geophys. Res.* **89**, 5929–5952 (1984).
7. Dziewonski, A. M. & Woodhouse, J. H. *Science* **236**, 37–48 (1987).
8. Cazenave, A., Souriau, A. & Dominh, K. *Nature* **340**, 54–57 (1989).
9. Cazenave, A., Dominh, K., Rabinowicz, M. & Ceuleneer, G. *J. geophys. Res.* **93**, 8064–8077 (1988).
10. Harrison, C. G. A., Miskell, K. J., Brass, G. W., Saltzman, E. S. & Sloan, J. L. *Tectonics* **2**, 357–377 (1983).
11. Algeo, T. J. & Wilkinson, B. H. *J. geol. Soc. London*. (in the press).
12. Harrison, C. G. A. *et al. Earth planet. Sci. Lett.* **54**, 1–16 (1981).
13. Anderson, D. L. *Nature* **297**, 391–393 (1982).
14. Gurnis, M. *Nature* **332**, 695–699 (1988).
15. Gurnis, M. *Geophys. Res. Lett.* (in the press).

ACKNOWLEDGEMENTS. This work was funded by a NSF Presidential Young Investigator Award. H. Pollack, M. Richards and B. Wilkinson provided helpful comments.

Iberian plate kinematics: a jumping plate boundary between Eurasia and Africa

S. P. Srivastava*, H. Schouten†, W. R. Roest*, K. D. Klitgord‡, L. C. Kovacs§, J. Verhoef* & R. Macnab*

* Geological Survey of Canada, Atlantic Geoscience Centre, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, B2Y 4A2, Canada

† Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

‡ US Geological Survey, Woods Hole, Massachusetts 02543, USA

§ US Naval Research Laboratory, Washington, DC, 20375-5000, USA

THE rotation of Iberia and its relation to the formation of the Pyrenees has been difficult to decipher because of the lack of detailed sea-floor spreading data, although several models have been proposed^{1–7}. Here we use detailed aeromagnetic measurements from the sea floor offshore of the Grand Banks of Newfoundland to show that Iberia moved as part of the African plate from late Cretaceous to mid-Eocene time, with a plate boundary extending westward from the Bay of Biscay. When motion along this boundary ceased, a boundary linking extension in the King's Trough to compression along the Pyrenees came into existence. Finally, since the late Oligocene, Iberia has been part of the Eurasian plate, with the boundary between Eurasia and Africa situated along the Azores–Gibraltar fracture zone.

The sea-floor spreading lineation pattern between the Azores and Charlie Gibbs fracture zones as obtained from the present survey and previous compilations^{8–12} (Fig. 1) provides the primary input into a reanalysis of the plate kinematics for the Iberian plate. The continuation of anomaly 34 into the Bay of Biscay¹³ indicates that Iberia was separating from Eurasia during the late Cretaceous. Analysis of sea-floor spreading anomaly patterns on the Iberian plate between the Azores–Gibraltar transform zone and King's Trough, in conjunction with the patterns south of the Azores–Gibraltar zone and on the west flank of the Mid-Atlantic Ridge^{8,11} showed that Iberia was attached to Africa from chron M0 to 13 (that is, 118 to 37 Myr). Further analysis of these data and new data from the present aeromagnetic survey east of the Grand Banks shows that the northern plate boundary of Iberia was more complicated than the single plate boundary at King's Trough used in the past^{6,8,11}. Examination of magnetic lineations from the west flank of the ridge which we rotated to the east using arbitrary poles of rotation (Fig. 2) indicates that the sea-floor spreading pattern on the eastern flank of the Mid-Atlantic Ridge can be divided

into four distinctive zones. These zones are separated by boundaries near the latitude of the Bay of Biscay, at King's Trough and at the present Azores-Gibraltar boundary. Here we focus on how the boundary shifted southwards through time.

A possible way to solve the misfit shown in Fig. 2 is to calculate independent rotation poles for the anomalies in each of the four zones. But the poles for zones 2 and 3 would not be well constrained because of the short length of the anomalies and the lack of well defined fracture zones. Olivet *et al.*⁵ used the Gloria fault, which is a part of the Azores-Gibraltar fracture zone, as a constraint for the independent motion of Iberia (zone 3). It is possible that this fracture zone always marked the Iberia-Africa boundary ever since Iberia separated from the Grand Banks. It is difficult, however, to use it as a constraint in deriving a satisfactory plate kinematic solution for Iberia, because of overprinting by more recent tectonic activity. Similarly, the geology of King's Trough has been interpreted differently by different authors^{4,5,14}, and cannot be used as a primary constraint for the calculation of poles of rotation.

Rather than proposing the existence of several small plates in the eastern North Atlantic, we have used the model of Schouten *et al.*⁶, in which Iberia has moved as part of much larger plates—Africa and Eurasia—separated by a jumping plate boundary. This has the advantage that the poles of rotation can be well constrained on account of the large sizes of the plates (several thousands as opposed to hundreds of kilometres) and the presence of numerous fracture zones. The problem then reduces to determining the location of the boundaries and the timing of the jumps.

The concept of a jumping plate boundary is not new. Based on very limited data, Smith¹⁵ was the first to propose a kinematic model for Iberia where it was intermittently attached to either Eurasia or Africa. The possibility that Iberia formed part of the African plate for a long period of time has been suggested by palaeomagnetic measurements^{16,17}, and the idea was further explored by Schouten *et al.*⁶, who examined the relative motion between Iberia and Eurasia in relation to the formation of King's Trough. They proposed that Iberia was part of Africa from early Cretaceous to Oligocene time and that the boundary between Eurasia and Africa extended from King's Trough to the Pyrenees before jumping to the Azores-Gibraltar fracture zone⁶. Here we expand on this model and suggest that the boundary along King's Trough was only active from middle Eocene to late Oligocene time. Before that, Iberia was attached to Africa, but the plate boundary was located farther north.

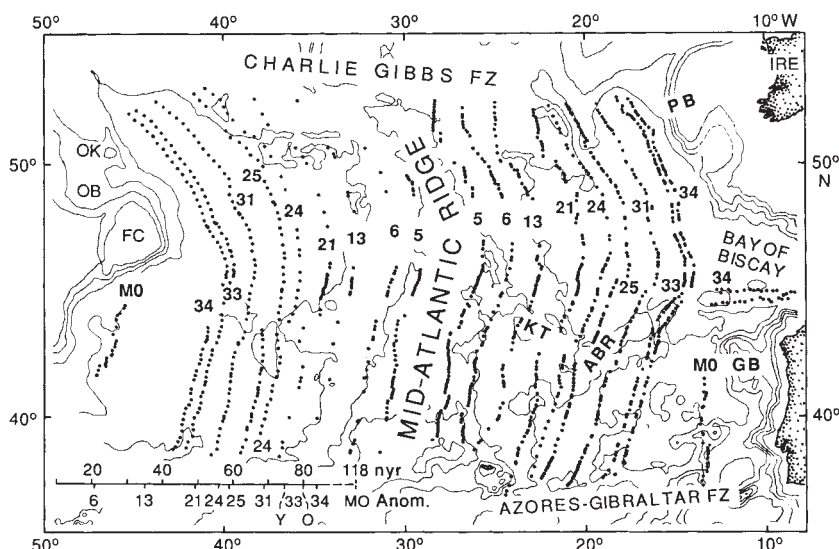
The geometry of a jumping plate boundary is shown schematically in Fig. 3. In this model it is assumed that Iberia

(IBR) and Africa (AFR) move together until the boundary between Iberia and Eurasia (EUR) jumps to the south of Iberia, and Iberia begins to move with Eurasia. The model shows how a shifting plate boundary can cause the shape of lineations on opposite sides of the ridge axis to be different, and how the anomalies on the eastern side may be divided into four zones. It is important to note that this model postulates motion among three rigid plates only, requiring only two independent poles of relative motion at any given time. We use this model to explain the observed differences between the lineations on the two sides of the North Atlantic (Fig. 2) and have derived a kinematic solution for the motion of Iberia which explains satisfactorily not only the formation of the Pyrenees, but also of King's Trough.

Figure 4 shows how the jumping-plate-boundary model can be applied to explain the observed sea-floor spreading anomalies. We will discuss the fit of anomalies proceeding from south to north. In zone 4 (Africa, anomalies not shown here) we have used the existing poles of rotation relative to North America^{11,18}. In zones 2 and 3, we first rotated the western anomalies to the east using the same North American/African reconstruction poles. For the sake of simplicity, we only show selected identifications in zone 3 to show that none of the rotated anomaly locations from the west (open circles) lies on the corresponding lineations in the east (lines). But with an additional clockwise rotation that corrects for the late Cenozoic motion along the Azores-Gibraltar fracture zone (pole P_1 , Fig. 4), portions of lineations 34 to 21, and possibly 13, can be matched in zone 3. This match can be extended to zone 2 by application of a second clockwise rotation (pole P_2 , Fig. 4). In zone 2, anomalies 34 to 21 which are rotated first with pole P_1 (open triangles) and then with pole P_2 (solid triangles), can now be matched as far as a boundary that extends west from the Bay of Biscay (labelled B in Fig. 4, corresponding to boundary x in Fig. 3). Application of further corrections to the anomalies in zone 1, north of boundary B does not improve the match, and here the North America/Eurasia poles^{12,19} were used. Boundary B is marked by a series of offsets in the eastern magnetic anomalies (Fig. 4).

The fact that anomalies 21 to 34, and possibly 13, in zones 2 and 3 can be matched with the African poles of rotation (with corrections) strongly suggests that these anomalies could have been formed on the African plate, and that the Africa-Eurasia boundary may have jumped successively from B to King's Trough to the Azores-Gibraltar fracture zone or from x to y to z in Fig. 3. From the match of anomalies (Fig. 4), and from calculations of differential poles of rotation between plates, we

FIG. 1 Identified sea-floor spreading magnetic anomalies in the North Atlantic between the Charlie Gibbs and the Azores-Gibraltar fracture zones, obtained from the present survey and previous compilations⁸⁻¹², superimposed on simplified bathymetric contours. FC, Flemish Cap; KT, King's Trough; ABR, Azores Biscay Rise; GB, Galicia Bank. Also shown are the ages of the magnetic anomalies²⁹.



find that, at about chron 19 (44 Myr ago) boundary B became extinct. The geological data from King's Trough¹⁴ indicate that from 35–25 Myr it was rifting apart suggesting that from about 44–25 Myr, a boundary extending from King's Trough to the Pyrenees could have been the main Africa–Eurasia boundary. On the other hand, because of the limited number of identifications of anomaly 13, the African connection of Iberia is not well established, and the data do not rule out independent motion at that time. The match of anomaly 6 (20 Myr) using the Eurasia/North America poles shows clearly that Iberia had begun to move with Eurasia by that time, with the Africa–Eurasia boundary in its present position along the Azores–Gibraltar fracture zone.

It is important to note that anomaly M0 does not match (Fig. 4), which indicates that Iberia at that time was not yet moving with Africa; this agrees with recent palaeomagnetic observations²⁰. It is also important to note that these reconstructions, using sea-floor spreading anomalies, are used to test a jumping-plate-boundary model; they do not rule out the possibility of simultaneous but mostly insignificant motion along other boundaries.

The model presented here describes relative motions along the shifting plate boundary that agree well with observations. Correction pole P_1 gives the total relative motion of Iberia relative to Africa along a complex and prominent series of bathymetric features—Azores–Gibraltar fracture zone¹¹. Our one plate-boundary-jumping model implies that this would be the total motion since 25 Myr. But if Iberia was acting as an independent plate after chron 19 then it would represent the total motion over a longer period of time, namely since the middle Eocene extinction of boundary B. This total motion agrees remarkably well with the observed extension at the Azores triple junction and with the timing, magnitude and direction of compression east of 19° W in the region of the Alboran back-arc basin²¹. The total motion, although strictly speaking not the present-day motion, also agrees quite well with the sense of

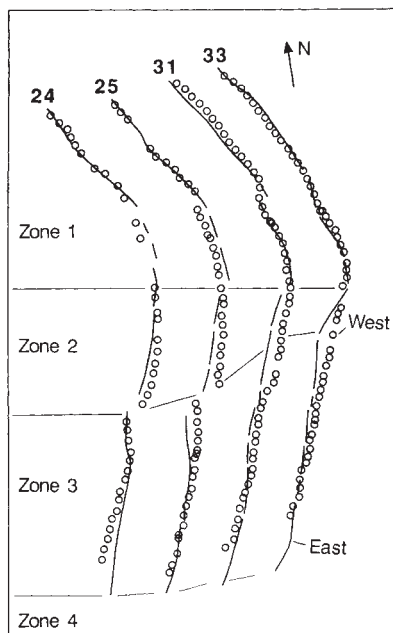


FIG. 2 The differences between the shapes of the western (circles) and eastern (lines) anomalies are shown. Anomalies 24 to 33, superimposed by rotation and plotted on a Lambert Conformal projection, are shown at arbitrary separations. As will be shown in Fig. 4, a much better fit is obtained if we divide the lineations into three zones. A fourth zone, south of the Azores–Gibraltar fracture zone, represents the anomalies on the African plate.

motion obtained from the focal mechanism solutions for present-day earthquakes along the Azores–Gibraltar fracture zone^{22,23} and with the strike-slip motion inferred along the Gloria Fault²². The implied motion between Africa and Iberia since the late Oligocene (or middle Eocene) agrees with the Oligocene–Miocene compression observed in the Iberian Betic and African Rift mountain system²¹.

Similarly, correction pole P_2 accounts for the total motion along the King's Trough/Azores–Biscay Rise/Pyrenees boundary from the middle Eocene extinction of boundary B to the end of activity in King's Trough in the late Oligocene, when Iberia began to move with Eurasia (Fig. 4). The location of this boundary was deduced from the fit of the magnetic anomalies, with incorporation of important bathymetric trends⁴. Figure 4

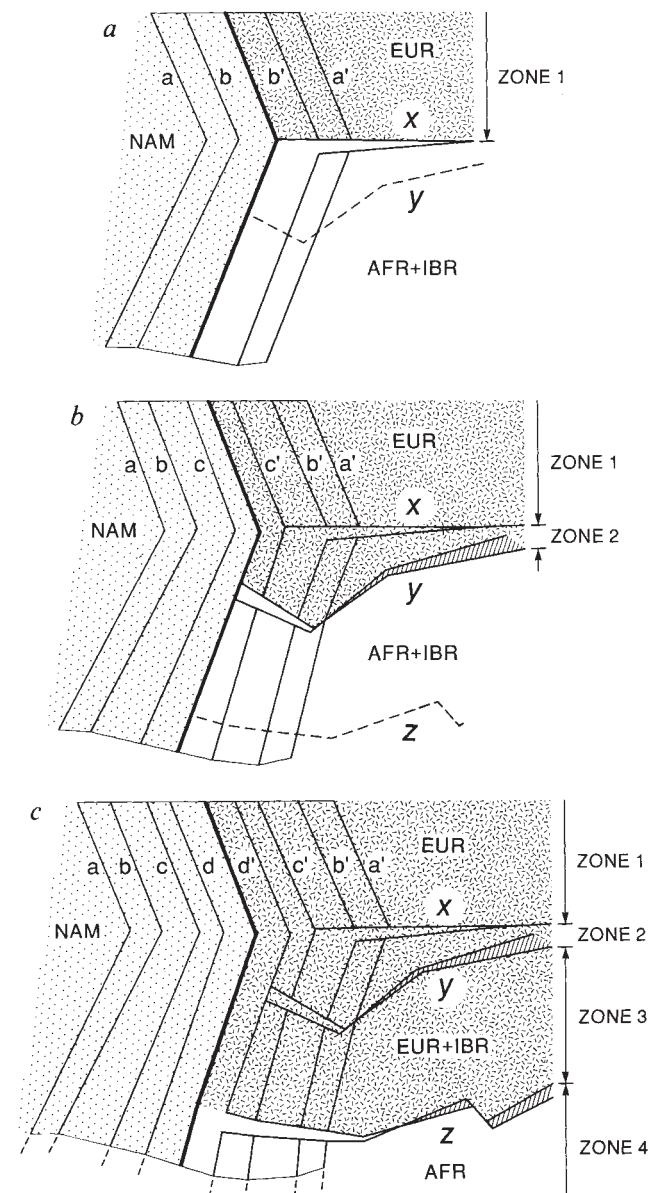


FIG. 3 Successive jumps of a plate boundary create differently shaped lineations on opposite sides of a ridge (compare the entire length of a and a' in panel a and c). In this model, boundary x between Eurasia (EUR) and Africa–Iberia (AFR + IBR) jumps first to location y. As a result, anomalies a' to c' formed in zone 2 as part of Africa started to move with Eurasia. Next, the boundary y jumps to location z, and Iberia becomes attached to Eurasia (IBR + EUR). As a result, all anomalies in zone 2 and 3 now move with Eurasia. NAM, North America.

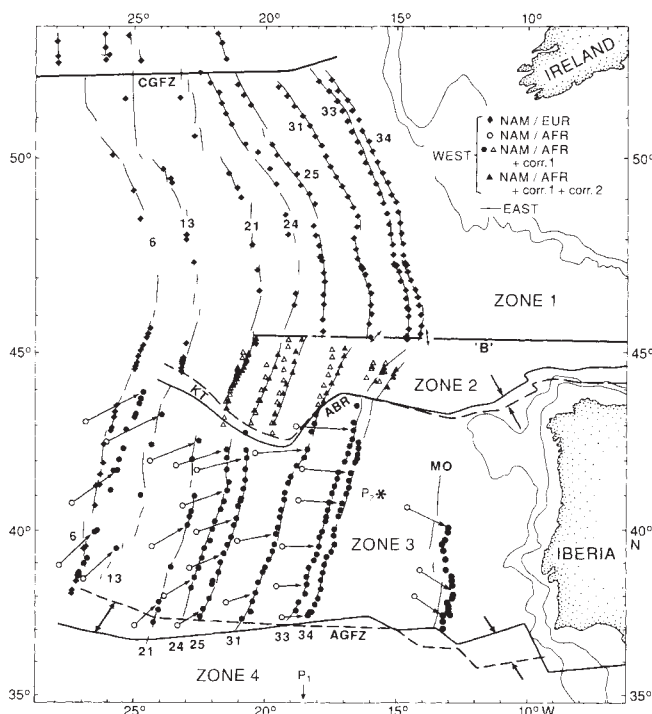


FIG. 4 Fit between western (symbols) and eastern (lines) lineations plotted on a Mercator projection. The open circles are western anomalies rotated to the east with North America/Africa poles of rotation^{11,18}. (For clarity, we show a few points in zone 3 only.) Open circles move to solid circles (zone 3) and to open triangles (zone 2) when a rotation of -7.87° about a pole P_1 at 31.43°N and 18.6°W , is applied. Similarly, application of an additional rotation of -4° about a pole P_2 at 40.8°N , 15.5°W in zone 2, rotates the open triangles to solid triangles. The two correction poles, P_1 and P_2 , represent the total motion that has taken place along the Azores-Gibraltar fracture zone (AGFZ) and the King's Trough/Azores-Biscay Rise (KT/ABR) boundary, as shown by the displacement between solid and dashed lines. Boundaries B, KT/ABR and AGFZ correspond to boundaries x, y and z of Fig. 3 respectively. Arrows inserted along boundary B show the instantaneous motion that took place along this boundary at the time of the isochron they are close to. The solid diamonds represent locations of western anomalies rotated to the east using North America/Eurasia poles of rotation. Boundary B was active from chron MO (118 Myr) to chron 19 (44 Myr), boundary KT/ABR from chron 19 to pre-chron 6 (~ 25 Myr) and boundary AGFZ since pre-chron 6 or possibly since chron 19.

shows that King's Trough was formed by extension across the plate boundary (compare the dashed and solid line). East of pole P_2 , compression took place along the north Spanish Trough and the Pyrenees.

Boundary B divided Africa and Eurasia for a long period of time, and different modes of motion existed across it. After the opening of the Bay of Biscay (Aptian to Campanian)²⁴, which was accompanied by compression along the Pyrenees²⁵, our model implies extensional motion along boundary B until chron 31. Between chrons 31 and 21, the motion along boundary B was mainly strike-slip, as evidenced by a gradual change in offset: from a dextral offset at anomaly 31, to a small sinistral offset at anomaly 21 (Fig. 4). No significant offset is observed in the anomalies west of the Mid-Atlantic Ridge at this latitude (Fig. 1). The strike-slip nature of this motion may explain the absence of major bathymetric features along the western part of this boundary.

The detailed sea-floor spreading data in the North Atlantic thus provide strong evidence that Iberia moved as part of the African plate from late Cretaceous to middle Eocene. From the middle Eocene to late Oligocene, it may also have moved as part of Africa or as an independent plate with only slight motion relative to the African plate, at least in the oceanic region. The relative motion between Eurasia and Iberia during this period was larger and resulted in the formation of King's Trough. Since the late Oligocene, Iberia has been moving as part of the Eurasian plate with its southern boundary along the Azores-Gibraltar fracture zone. Our model shows that before the late Oligocene the motion between Eurasia and Africa was mainly confined to a boundary north of Iberia. Thus, the relative motion between Eurasia and Africa postulated by our model can provide direct constraints on Alpine tectonism up to the late Oligocene. When the plate boundary shifted south to the Azores-Gibraltar fracture zone, a number of microplates were created in the Mediterranean²¹. The fragmentation of the northern Eurasia-Africa plate boundary during this time may thus be indicative of true continent-continent collision throughout the Mediterranean.

We do not claim our model to be the only solution for the motion of Iberia, but it is simple and provides a direct linkage

between tectonic events offshore and those on land. Furthermore, our model of a jumping plate boundary can serve as an attractive working model to interpret the kinematics of other microplate systems (for example, the Mesozoic and Cenozoic evolution of the Middle Americas^{26,27}, and the Pacific²⁸). □

Received 18 September 1989; accepted 26 February 1990.

1. Le Pichon, X. & Sibuet, J. C. *Earth planet. Sci. Lett.* **12**, 83-88 (1971).
2. Le Pichon, X., Sibuet, J. C. & Francheteau, J. *Tectonophysics* **38**, 169-209 (1977).
3. Slater, J. G., Hellinger, S. & Tapscott, C. R. *J. Geol.* **85**, 509-552 (1977).
4. Grimaud, S. et al. *Mar. Geol.* **45**, 63-77 (1982); *Mar. Geol.* **53**, 237-239 (1983).
5. Olivet, J. L., Bonnin, J., Beuzart, P. & Auzende, J. M. in *Publication du Centre National pour l'Exploitation des Océans*, Paris, 1-108 (Centre Océanologique de Bretagne, Brest, 1984).
6. Schouten, H., Srivastava, S. P. & Klitgord, K. *Trans. Am. geophys. Un.* **65**, 190 (1984).
7. Savostin, L. A., Sibuet, J. C., Zonenshain, L. P., LePichon, X. & Roulet, M. J. *Tectonophysics* **123**, 1-35 (1986).
8. Srivastava, S. P. & Tapscott, C. R. in *The Geology of North America* Vol. M (eds Vogt, P. R. & Tucholke, B. E.) 379-404 (Geol. Soc. Am., 1986).
9. Roberts, D. G., Jones, M. T. & Hunter, P. *Inst. Ocean Sci. Rep.* **207** (Wormley, 1985).
10. Verhoef, J. et al. *Mar. geophys. Res.* **8**, 1-25 (1986).
11. Klitgord, K. & Schouten, H. in *The Geology of North America* Vol. M (eds Vogt, P. R. & Tucholke, B. E.) 351-378 (Geol. Soc. Am., 1986).
12. Srivastava, S. P., Verhoef, J. & Macnab, R. *Mar. Petrol. Geol.* **5**, 306-337 (1988).
13. Kristoffersen, Y. *Earth planet. Sci. Lett.* **38**, 273-290 (1978).
14. Kidd, R. & Ramsay, A. T. S. *Init. Rep. DSDP*, Leg. **94**, 1245-1261 (1986).
15. Smith, A. *Geol. Soc. Am. Bull.* **82**, 2039-2070 (1971).
16. Van der Voo, R. *Tectonophysics* **7**, 5-56 (1969).
17. Van den Berg, J. & Zijdeveld, J. D. A. *AGU Geodynamics Ser.* **7**, 83-112 (1982).
18. Roest, W. R. thesis, Univ. Utrecht (1987).
19. Srivastava, S. P. & Roest, W. R. *East Coast Basin Atlas Series: Labrador; Atlantic Geoscience Centre, Geol. Surv. Canada* 100-109 (Dartmouth, 1989).
20. Galdeano, A., Moreau, M. G., Pozzi, J. P., Berthou, P. Y. & Malod, J. A. *Earth planet. Sci. Lett.* **92**, 95-106 (1989).
21. Horvath, F. & Berckhemer, H. *AGU Geodynamics Ser.* **7**, 141-173 (1982).
22. Argus, D. F., Gordon, R. G., DeMets, C. & Stein, S. J. *geophys. Res.* **94**, 5585-5602 (1989).
23. McKenzie, D. P. *Geophys. J. R. astr. Soc.* **30**, 109-185 (1972).
24. Montadert, L., Roberts, D. G., de Charpal, O. & Guennoc, P. *Init. Rep. DSDP*, Leg. **48**, 1025-1060 (1979).
25. Mattauer, M. & Henry, J. *Publ. geol. Soc. Lond.* **4**, 3-22 (1974).
26. Pindell, J. L. & Dewey, J. F. *Tectonics* **1**, 179-211 (1982).
27. Burke, K., Cooper, C., Dewey, J. F., Mann, P. & Pindell, J. L. *Geol. Soc. Am. Mem.* **162**, 31-63 (1984).
28. Atwater, T. *The Geology of North America* (Geol. Soc. Am., in the press).
29. Kent, D. V. & Gradstein, F. M. in *The Geology of North America* Vol. M (eds Vogt, P. R. & Tucholke, B. E.) 45-50 (Geol. Soc. Am., 1986).

ACKNOWLEDGEMENTS. We thank Mike Keen and Matthew Salisbury for their comments on the manuscript and many of our colleagues from the Atlantic Geoscience Centre for their assistance during the course of this work. This study was carried out under the Frontier Geoscience Program of the Geological Survey of Canada. W.R. is supported by the Natural Sciences and Engineering Research Council of Canada, under a Visiting Fellowship.

Regional Al-Fe mafic magmas associated with anorthosite-bearing terranes

K. E. Olson & S. A. Morse

Department of Geology and Geography, University of Massachusetts, Amherst, Massachusetts 01003, USA

PROTEROZOIC anorthosite massifs are nearly monomineralic accumulations of plagioclase feldspar, which must result from unusual conditions of crystallization and emplacement^{1,2}. Their chemistry is compatible with derivation from a high-alumina-basalt parent of mantle origin¹, and mafic magmas rich in aluminium and iron have long been suspected of having an important connection to anorthosites and their parent magmas¹⁻⁴. Regional examples of mafic magmas clearly compatible with an anorthosite lineage have not been generally recognized³, in part because of the paucity of candidates within the anorthosite massifs themselves. Samples of such magmas should also be sought in the country rocks surrounding the massifs. Here we report the widespread occurrence of Al-Fe mafic magmas in the anorthosite-bearing Adirondack Highlands of New York state. They are chemically distinct from typical high-Al basalt, Fe basalt and most common magma types. Identification of similar Al-Fe basaltic magmas in other areas suggests that such magmas are characteristic of anorthosite-bearing terranes.

Mafic igneous rocks in the 20,000 km² Adirondack region of New York state (Fig. 1) make up ~8% of the surface exposure of the area. The mafic suite⁵ includes metamorphosed gabbroic rocks and gabbroic troctolites emplaced anorogenically with anorthosite at about 1,150 Myr, between pulses of the Grenville orogenic event⁶. A mantle source is assumed for these rocks because mafic (broadly basaltic) magmatism is best explained by adiabatic decompression of hot mantle rock.

Figure 2 summarizes the major-element concentrations in Adirondack gabbroic rocks. The rocks have the unusual combination of low Si relative to common tholeiites and high Al and

Fe comparable with typical high-Al basalts and Fe basalts. High Al resembles orogenic high-Al basalts from western Japan⁷ and the Medicine Lake Highlands, California⁸⁻¹⁰, which, however, show Si rather than Fe enrichment (Bowen as opposed to Fenner trends). Derivation of high-Al basalt parental to anorthosite has been shown to occur by delayed nucleation of feldspar in hyperfeldspathic magmas²⁻⁴. For example, high-Al and low-Al gabbros in the anorthosite-related Harp Lake complex of Labrador can be related by the fractionation of plagioclase, which accumulated with residual liquid as anorthosite⁴. A similar process may relate Adirondack high-Al gabbro and anorthosite.

Magmas with FeO* > 12 weight % are Fe basalts as defined by Archaean volcanic rocks from the Superior Province^{11,12} and by modern Fe-Ti basalts from Galapagos¹³⁻¹⁵. Most of the Adirondack samples fall into this class, yet they have higher Al contents than typical Fe basalts and Fe-Ti basalts. High Fe in the Adirondack rocks is due in part to Fe enrichment during fractionation reflecting a low oxidation state of the magmas as typified by the Skaergaard Intrusion, and in the extreme case by the Kiglapait Intrusion^{16,17}. High Fe is a feature of anorogenic magmas and anorthosite terranes²⁻⁴ and reflects intrinsic Fe richness of the parental magmas. Fractional crystallization alone cannot account for the observed Fe enrichment because it must invariably lead to augite saturation, whereas the augite component in these magmas is low². The presence of elevated Fe in the source rocks is compatible with models that derive anorthosite from old enriched mantle², and the suspected higher overall Fe and Al in the ancient mantle¹⁸. The Fe-rich signature of Adirondack gabbros in the presence of low Si and high Al is therefore the product of source chemistry combined with fractionation history.

Trace elements are illustrated by the rare-earth elements (REE) in Fig. 3. The Adirondack rocks are characterized by light REE (LREE) enrichment for a range in element abundances, reflecting source enrichment² or contamination of the magmas¹⁹. The REE patterns of Al-Fe rocks are geochemically distinct from Al- or Fe-rich modern and Archaean basalts (Fig. 3). Flat slopes of Archaean Fe-basalt and steep slopes of Archaean Al-basalt differ from those of the Adirondack rocks. Fe-Ti basalts from the Galapagos show the flat to LREE-

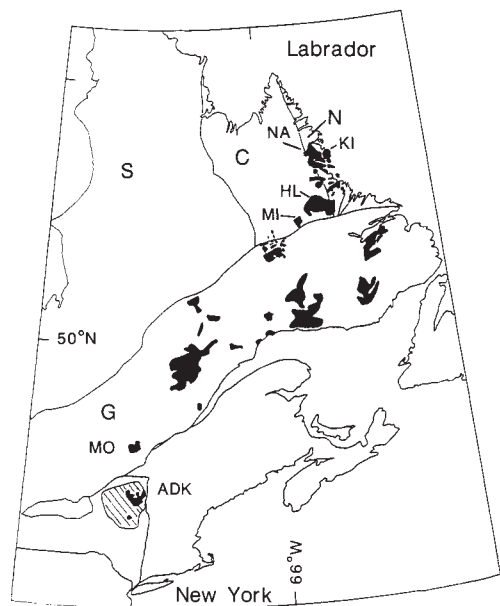


FIG. 1 The eastern Canadian shield, New York to Labrador, with the major anorthosite massifs (NA: Nain; HL: Harp Lake; MI: Michikamau; MO: Morin; ADK: Adirondack), Kiglapait Intrusion (KI), and structural provinces (N: Nain; C: Churchill; S: Superior; G: Grenville). Hatched pattern shows the study area, the Adirondack region.

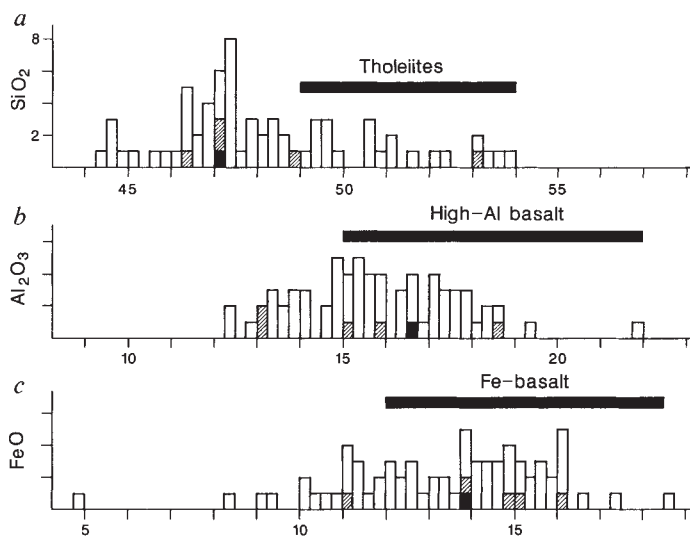


FIG. 2 Histograms for SiO₂, Al₂O₃ and FeO* (total Fe as FeO) in 67 Adirondack mafic rocks. Bars show range of SiO₂ in tholeiites²³, Al₂O₃ in high-Al basalt⁷, and FeO* in Fe-basalt¹¹. Striped and solid blocks represent gabbros and gabbroic troctolite that constitute the Adirondack field in Fig. 3.

depleted patterns and lower overall abundances characteristic of low-pressure fractionation in depleted mantle-derived ocean-floor basalts. Modern orogenic Al-basalts are variably LREE enriched, reflecting complex histories involving mixing in arc environments.

The Adirondack gabbroic rocks are the first regional examples of Al-Fe magmas to be recognized. Chilled margins at dry contacts confirm the local existence of such magmas near anorthosite massifs. They have been described in the Kiglapait and Hettasch Intrusions in Labrador, where Al-rich, SiO_2 -undersaturated magmas showing Fe enrichment occur at the margins of the Nain anorthosite²⁰. The combination of Fe enrichment and extended crystallization of plagioclase requires the presence of large volumes of such magmas over an extended time interval. We feel these magmas are represented by the Adirondack rocks.

In Fig. 4a, the Adirondack mafic suite ranges from plagioclase-rich to olivine-rich normative compositions through a mid-point that lies in the upper right quadrant of the diagram (asterisk, Fig. 4b, c). The presence of a gap between the mafic members of the suite and the anorthosites, combined with the absence of endmember ultramafic magmas, suggests that the observed spread is not a relic of endmember magma mixing. Instead, the Adirondack data reflect the relative accumulation of plagioclase and olivine (or orthopyroxene) from a composition initially high in Al and Fe such as gabbroic troctolite. The distribution of the

Adirondack rocks in Fig. 4a defines the region as a 20,000 km² Al-Fe petrographic province (Fig. 1).

Mafic magmas from other anorthosite-bearing regions follow the same trend as the Adirondack rocks. Figure 4b shows the data from Labrador, and other regions are shown in Fig. 4c. The calculated liquid compositions of the Kiglapait Intrusion (Fig. 4b, heavy curve) illustrate the case for closed-system fractionation of an Al-Fe magma associated with anorthosite. Fields for other Labrador examples also follow this trend. The Sierra Ancha sill complex (Fig. 4c) contains microtroctolite and anorthosite, and is an analogue of the Kiglapait Intrusion. Leuconorites from Morin, and high-Al (low- and intermediate-K) Fra Mauro basalts associated with the anorthositic lunar highlands follow a similar trend.

Figure 4d shows for comparison high-Al basalts from western Japan and California, Fe-Ti basalts from the Galapagos, and representative ocean floor and Hawaiian basalts. Comparison of the diagrams in Fig. 4 reveals the control by clinopyroxene (plotted in Fig. 4a) in the chemistry of modern Al- and Fe-rich basalts (Fig. 4d), and its absence in the augite-undersaturated Al-Fe magmas (Fig. 4b, c). The distinctiveness of the anorthosite-related Al-Fe magmas is therefore a reflection of bulk composition, highlighted by fractionation and represented by modal mineralogy.

The Al-Fe magma trend is unique, as predicted by Morse²

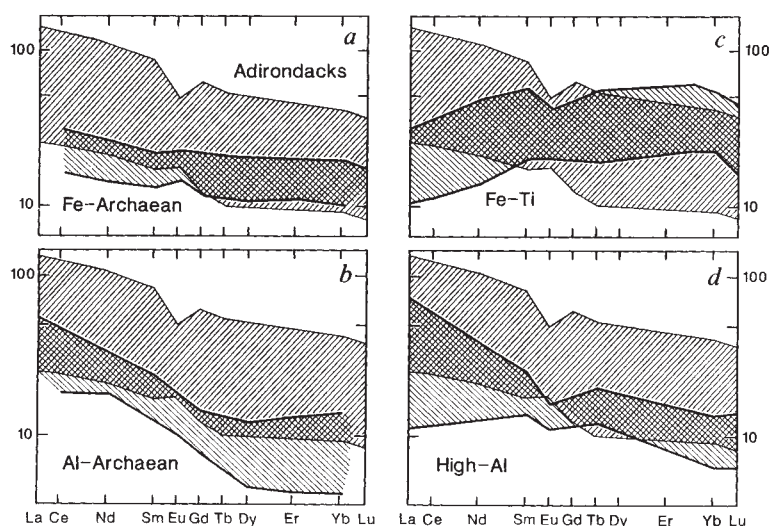
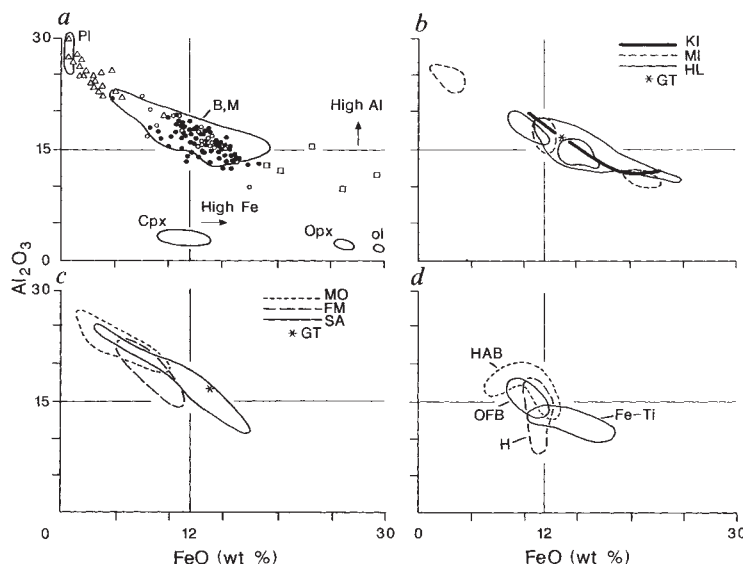


FIG. 3 Chondrite-normalized²⁴ rare-earth element patterns in Adirondack gabbroic rocks (repeated field) compared with *a*, Archaean Fe-basalt¹²; *b*, Archaean Al-basalt²⁵; *c*, ocean floor Fe-Ti basalt¹⁴; and *d*, orogenic high-Al basalt⁹.

FIG. 4 Plot of Al_2O_3 against FeO^* for *a*, Adirondack Al-Fe magmas; *b*, Labrador Al-Fe magmas; *c*, other Al-Fe magmas; and *d*, modern basalts. Reference lines show lower limits for high-Al basalt and Fe-basalt. *a*, Solid circles, gabbroic rocks; open circles, gabbroic troctolites; triangles, anorthosites; squares, mafic gabbros²⁶; B,M, published analyses^{27,28}, Adirondack²⁹ clinopyroxene (cpx) and orthopyroxene (opx); Kiglapait¹⁶ plagioclase (pl) and olivine (ol). *b*, Kiglapait liquid path¹⁷; Michikamau³⁰ anorthosite, gabbro and ferrodiorite; Harp Lake⁴ high-Al gabbro, low-Al gabbro and ferrodiorite. *c*, Morin³¹ leuconorites; lunar³² low-K and intermediate-K Fra Mauro basalts; Sierra Ancha³³ dolerite, microtroctolite and anorthosite. *d*, HAB, high-Al basalt, Japan and Medicine Lake Highlands⁷⁻¹⁰; OFB, ocean floor basalt³²; H, Hawaiian basalt³²; FeTi basalt, Galapagos¹³⁻¹⁵. Asterisk represents average Adirondack gabbroic troctolite, a possible parent composition.



and by Emslie³. Its association with anorthosite was recognized early in western Norway, where Eskola²¹ observed Al-Fe rocks whose best analogues were gabbros of the Adirondacks. The regional extent of the Adirondack rocks provides support to models producing anorthosites by proto-rift-related¹ or plume-generated and channelled magma pulses that may correspond to the widespread plutonism characteristic of anorogenic events in supercontinents²². Examples should be sought in other anorthosite-bearing terranes. □

Received 1 November 1989; accepted 20 February 1990.

1. Emslie, R. F. *Precamb. Res.* **7**, 61–98 (1978).
2. Morse, S. A. *Am. Miner.* **67**, 1087–1100 (1982).
3. Emslie, R. F. *Can. J. Earth Sci.* **15**, 438–453 (1978).
4. Emslie, R. F. *Bull. geol. Surv. Can.* **293**, 136 pp. (1980).
5. Olson, K. E. in *Spec. Pap. geol. Ass. Can.* (eds Gower, C. F., Rivers, T. & Ryan, B.) (in the press).
6. McLelland, J., Chiarenzelli, J., Whitney, P. & Isachsen, Y. *Geology* **16**, 920–924 (1988).
7. Kuno, H. *J. Petr.* **1**, 121–145 (1960).
8. Powers, H. A. *Am. Miner.* **17**, 253–294 (1932).
9. Condie, K. C. & Hayslip, D. L. *Geochim. cosmochim. Acta* **39**, 1165–1178 (1975).
10. Grove, T. L., Gerlach, D. C. & Sando, T. W. *Contr. Miner. Petrol.* **80**, 160–182 (1982).
11. Pearce, T. H. & Birkett, T. C. *Can. Miner.* **12**, 509–519 (1974).
12. Shirley, S. B. & Hanson, G. N. *Geochim. cosmochim. Acta* **50**, 2631–2651 (1986).
13. Christie, D. M. & Sinton, J. M. *Earth planet. Sci. Lett.* **56**, 321–335 (1981).
14. Schilling, J.-G., Kingsley, R. H. & Devine, J. D. *J. geophys. Res.* **87**, 5593–5610 (1982).
15. Juster, T. C., Grove, T. L. & Perfit, M. R. *J. geophys. Res.* **94**, 9251–9274 (1989).
16. Morse, S. A. *Mem. geol. Soc. Am.* **112**, (1969).
17. Morse, S. A. *Geochim. cosmochim. Acta* **45**, 461–479 (1981).
18. Jordan, T. H. *J. Petrol. Spec. Lithosph. Issue*, 11–37 (1988).
19. Ashwal, L. D. & Wooden, J. L. *Geochim. cosmochim. Acta* **47**, 1875–1885 (1983).
20. Nolan, K. M. & Morse, S. A. *Geochim. cosmochim. Acta* **50**, 2381–2386 (1986).
21. Eskola, P. *Norw. Vidensk. Skr. I. Mat.-Nat. Kl. no. 8*, p. 44 (1921).
22. Hoffman, P. F. *Geology* **17**, 135–138 (1989).
23. Yoder, H. S. & Tilley, C. E. *J. Petrol.* **3**, 342–532 (1962).
24. Haskin, L. A., Haskin, M. A. & Frey, F. A. in *Origin and Distribution of the Elements* (ed. Ahrens, L. H.) 889–912 (Pergamon, 1968).
25. Hawkesworth, C. J. & O'Nions, R. K. *J. Petrol.* **18**, 487–520 (1977).
26. Ashwal, L. D. thesis, Princeton Univ. (1979).
27. Buddington, A. F. *Am. J. Sci.*, *Bowen volume*, 37–84 (1952).
28. McLelland, J., Lochhead, A. & Vynhal, C. *J. Geol.* **96**, 279–298 (1988).
29. Ollila, P. W. thesis, Univ. Massachusetts (1984).
30. Emslie, R. F. *Can. J. Earth Sci.* **2**, 385–399 (1965).
31. Martignole, J. thesis, Univ. Montreal (1973).
32. Basaltic Volcanism Study Project *Basaltic Volcanism on the Terrestrial Planets* (Pergamon, 1981).
33. Nehru, C. E. & Prinz, M. *Geol. Soc. Am. Bull.* **81**, 1733–1766 (1970).

Denitrification by sulphur oxidizing *Beggiatoa* spp. mats on freshwater sediments

Jean-Pierre R. A. Sweerts*, Dirk De Beer†, Lars Peter Nielsen‡, Henk Verdouw*, Johannes C. Van den Heuvel†, Yehuda Cohen§ & Thomas E. Cappenberg*

* Limnological Institute, Rijkswaterstaatweg 6, 3631 AC Nieuwersluis, The Netherlands

† Department of Chemical Technology, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WW Amsterdam, The Netherlands

‡ Division for Freshwater Ecology, Danish Environmental Research Institute, Lysbrogade 52, 8600 Silkeborg, Denmark

§ H. Steinetz Marine Biology Laboratory, Eilat 88103, Israel

AT the surface of sulphide-rich marine and freshwater sediments, the gliding filamentous sulphur bacteria *Beggiatoa* spp. often grow abundantly to form conspicuous white mats^{1–6}. *Beggiatoa* may be responsible for the whole benthic consumption of oxygen, converting sulphide to either intracellular elemental sulphur, or to extracellular sulphate^{7–9}. Various strains of *Beggiatoa* have shown a capacity for nitrogen fixation and assimilation of ammonium and nitrate¹⁰. We now report that denitrification activity may also be attributed to *Beggiatoa* spp., which indicates that *Beggiatoa* mats could have a key role in the benthic cycling of oxygen, sulphur and nitrogen. Microprofiles of nitrate through a *Beggiatoa* mat on a lake sediment showed that all nitrate taken up from the water

TABLE 1 Characteristics of oxygen and nitrate microprofiles and calculated nitrate fluxes in sediments with and without *Beggiatoa* mats

Sample	[NO ₃] in water phase (μM)	O ₂ penetration (μm)	NO ₃ [−] penetration (μm)	NO ₃ [−] flux (mmol m ^{−2} day ^{−1})
No <i>Beggiatoa</i>	28	1,200	1,700	1.3
<i>Beggiatoa</i> mat	28	50	150	2.5
<i>Beggiatoa</i> mat	87	50	150	7.2

phase was consumed within the upper 150 μm of the mat. Complete denitrification was demonstrated by the conversion of ¹⁵NO₃[−] to ¹⁵N₂ in semi-purified tufts of *Beggiatoa*, and the coupling of denitrification and sulphide oxidation was indicated by microprofiles of oxygen and sulphide in pure cultures.

Mats of *Beggiatoa* cover ~25% of the deep profundal bottom of meso-eutrophic Lake Vechten¹¹ in the non-stratified period. Microscopic examination indicates that no bacteria other than *Beggiatoa* appear in significant numbers in the mat and, from the width of the filaments (3–4 μm), the population has been identified as *Beggiatoa alba*¹².

Using recently developed microsensors^{13–15}, we showed from oxygen and nitrate microprofiles that oxygen from the water phase was consumed completely in the upper 50 μm of the mat, whereas nitrate was consumed anaerobically in the next 100 μm (Fig. 1). Thus, no oxygen or nitrate penetrated the sediment below the 600 μm-thick *Beggiatoa* mat. The flux of nitrate was calculated from the linear concentration gradient in the diffusive boundary layer above the mat using Fick's first law of diffusion: $J = D_0 \cdot dC/dx$, where J is the flux, D_0 is the molecular diffusion coefficient for nitrate ($1.22 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ at 7 °C; ref. 16), and dC/dx is the slope of the concentration gradient. The nitrate consumption by the *Beggiatoa* was thus calculated to be 2.5 mmol m^{−2} day^{−1} (Table 1). We attribute this to denitrification, which gives a consumption rate that is among the highest reported for lake sediments¹⁷. In nearby sediment without *Beggiatoa* mats, denitrification occurred below 1.2 mm into the

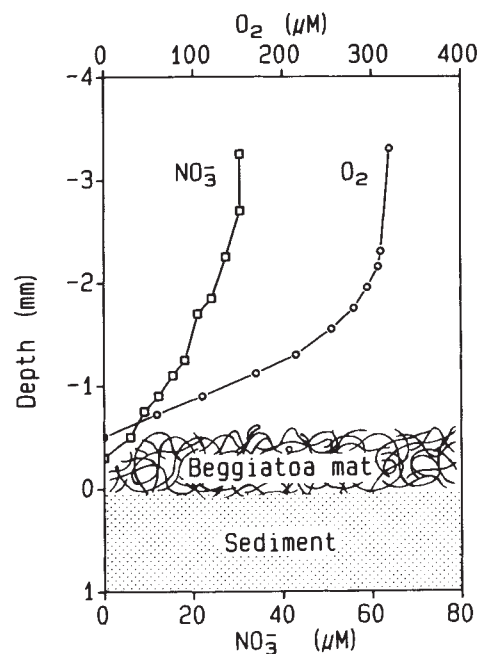


FIG. 1 Microprofiles of nitrate and oxygen in a mat of *Beggiatoa* spp. on lake sediment. (Freshly collected lake-sediment cores maintained at 7 °C with gentle stirring of the overlying water phase.)

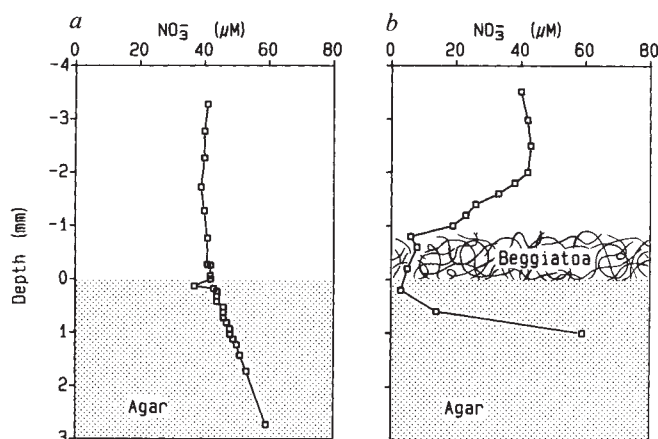


FIG. 2 Microprofiles of nitrate in an agar-water interface without bacteria (a) and with a semi-purified tuft of *Beggiatoa* (b). Agar composition: 0.5% agar, 0.2 mM acetate, 0.1 mM nitrate, 0.1 mM sulphide and 0.5 mM ammonium.

sediment, and the nitrate flux was only $1.3 \text{ mmol m}^{-2} \text{ day}^{-1}$ (Table 1).

As nitrate was consumed just below the mat-water interface (Fig. 1), the permeability of the diffusive boundary layer and the nitrate concentration in the water phase were both likely to control the nitrate flux, and hence the denitrification rate, in the mat. This was confirmed by our observation that if the nitrate concentration in the water phase was increased threefold, then the nitrate flux increased by roughly the same factor (Table 1).

The denitrification activity in the *Beggiatoa* mat could be ascribed to the few other bacteria present. Therefore tufts of *Beggiatoa* filaments were carefully rinsed with sterile filtered (Nucleopore, $0.2 \mu\text{m}$) lake water until microscopic examination showed that virtually all other bacteria had been removed. The tufts were then placed on agar supplied with nitrate; pronounced nitrate gradients developed within an hour (Fig. 2), indicating that denitrification was actively occurring as in the mat *in situ* (Fig. 1).

We added $^{15}\text{NO}_3^-$ to a final concentration of $100 \mu\text{M}$ to the water phase overlying the agar and *Beggiatoa* tufts, and monitored the release of $^{15}\text{N}_2$ gas to the overlying closed gas phase using a gas chromatograph/mass selective detector (GC-MSD; gas separation with a Poraplot-Q column; gas chromatograph, HP 5890 A; He carrier gas; 30°C ; $20 \mu\text{l}$ sample; mass spectrometer, MSD HP 5970 B). Within 48 h, 80% of the $^{15}\text{NO}_3^-$ added was recovered as $^{15}\text{N}_2$, demonstrating that nitrate was

consumed by denitrification and not by assimilation or dissimilatory reduction to ammonium.

The interaction of denitrification and sulphide oxidation by *Beggiatoa* was investigated by microsensing pure cultures of *Beggiatoa alba*, strain B18LD⁷, a well studied, mixotrophic, freshwater strain designated as the type strain of *Beggiatoa alba*². Growing in fluid medium causes the strain to form round tufts; we mounted single tufts on glass needles in a sulphide-free medium and measured the effect of nitrate on the concentration profiles of oxygen and sulphide in the tuft with microsensors¹⁵ (Fig. 3). In the central anoxic part of the tuft the intracellular pools of elemental sulphur were reduced to sulphide, which diffused towards the periphery where it was reoxidized. Without nitrate, oxygen diffusing in from the outside met with sulphide in a zone $50 \mu\text{m}$ wide, in which all sulphide was apparently oxidized aerobically (Fig. 3a). When nitrate was added, a $100 \mu\text{m}$ -wide gap appeared between oxygen and sulphide, showing that sulphide was now being consumed anaerobically, obviously by oxidation with nitrate.

The results clearly illustrate the denitrifying capacity of the investigated freshwater *Beggiatoa* spp. The formation of mat structures on sediment surfaces along with a high metabolic capacity seems to reflect an optimal adaptation of *Beggiatoa* spp. to the utilization of both oxygen and nitrate diffusing from the overlying water phase. The presence of *Beggiatoa* mats thereby increases nitrogen removal in lake systems. The bottom of many coastal seas and upwelling areas are also covered with mats of species of *Beggiatoa* or of the genus *Thioploca*, and if these marine species are also denitrifiers, then they may be a cause of the nitrogen limitation of phytoplankton growth in the sea. □

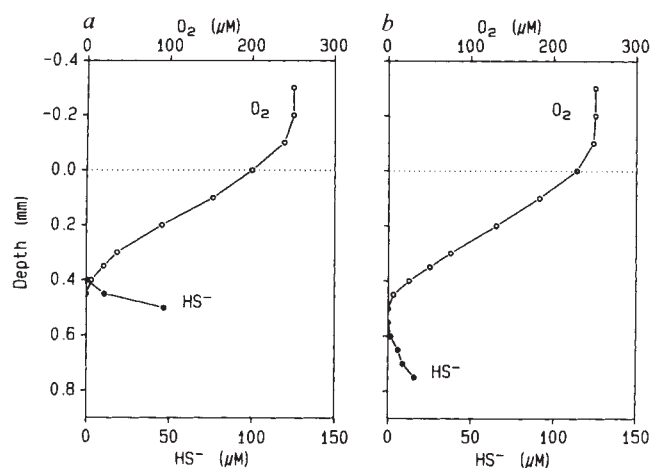


FIG. 3 Microprofiles of oxygen and sulphide in a 3-mm-broad tuft of *Beggiatoa alba*, strain B18LD, in the absence of (a) and in the presence of (b) nitrate ($300 \mu\text{M}$) (medium with 3.6 mM acetate and no sulphide).

Received 6 November 1989; accepted 12 February 1990.

- Winogradsky, S. *Bot. Zg.* **45**, 489–610 (1887).
- Larkin, J. M. & Strohl, W. R. *A. Rev. Microbiol.* **37**, 341–367 (1983).
- Jørgensen, B. B. *Mar. Biol.* **41**, 19–28 (1977).
- Smith, C. R., Kukert, H., Wheatcroft, R. A., Jumars, P. A. & Deming, J. W. *Nature* **341**, 27–28 (1989).
- Lackey, J. B. *Water & Sewage Works* **108**, 29–31 (1961).
- Jørgensen, B. B. & Revsbech, N. P. *Appl. environ. Microbiol.* **45**, 1261–1270 (1983).
- Strohl, W. R. & Larkin, J. M. *Appl. environ. Microbiol.* **36**, 755–770 (1978).
- Strohl, W. R., Geffers, I. & Larkin, J. M. *Curr. Microbiol.* **6**, 75–79 (1981).
- Nelson, D. C. & Castenholz, R. W. *J. Bact.* **147**, 140–154 (1981).
- Nelson, D. C., Waterbury, J. B. & Jannasch, H. W. *Arch. Microbiol.* **133**, 172–177 (1982).
- Steenbergen, C. L. M. & Verdouw, H. *Hydrobiologia* **95**, 11–23 (1982).
- Leadbetter, E. R. in *Bergey's Manual of Determinative Bacteriology* (eds Buchanan, R. E. & Gibbons, N. E.) 8th edn, 112–116 (Williams and Wilkins, Baltimore, 1974).
- De Beer, D. & Sweets, J.-P. R. A. *Analyt. chim. Acta* **219**, 351–356 (1989).
- Sweets, J.-P. R. A. & De Beer, D. *Appl. environ. Microbiol.* **55**, 754–757 (1989).
- Revsbech, N. P. & Jørgensen, B. B. *Adv. microb. Ecol.* **9**, 293–352 (1986).
- Li, Y.-H. & Gregory, S. *Geochim. cosmochim. Acta* **38**, 703–714 (1974).
- Seitzinger, S. P. *Limnol. Oceanogr.* **33**, 702–724 (1988).

ACKNOWLEDGEMENTS. We thank M. J. Bär-Gillisen for technical assistance.

Genetic fingerprinting reflects population differentiation in the California Channel Island fox

Dennis A. Gilbert*, Niles Lehman†, Stephen J. O'Brien‡ & Robert K. Wayne†§

* Biological Carcinogenesis and Development Program, Program Resources Incorporated, NCI-FCRF, Frederick, Maryland Z1701, USA, and The Johns Hopkins University, Department of Biology, Baltimore, Maryland 21218, USA

† Department of Biology, UCLA, Los Angeles, California 90024, USA

‡ Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, Maryland Z1701, USA

RESTRICTION fragment profiles generated by hybridization of hypervariable minisatellite DNA probes have been used for paternity analysis but not for comparisons at the level of populations, because the profiles are thought to evolve too rapidly to be informative over large time intervals¹⁻³. But in small isolated populations, the fixation of restriction-fragment polymorphisms can outpace the generation of fragment-length variability through recombination. Here we report on an analysis of DNA fingerprints of the California Channel Island fox (*Urocyon littoralis*). These foxes comprise an island dwarf species found only on six of the Channel Islands off the coast of southern California^{4,5}. Variability of restriction-fragment profiles within fox populations, as indicated by the average percentage difference (APD), varied widely among the islands, from 0.0% (no variation) to 25.3%. The APDs between populations were considerably greater (43.8% to 84.4%). In addition, foxes on each island can be distinguished by the presence of diagnostic restriction fragments. Maximum parsimony and phenetic trees relating foxes from different islands are consistent with the archaeozoological and geological record. Therefore, in small populations of genetically isolated mammals, differences among hypervariable restriction-fragment profiles can be used to estimate relative genetic variability and to reconstruct the evolutionary relationships of natural populations.

For a map showing the six Channel Islands on which foxes are found, and the area, population size, and time of establishment of each island, see Fig. 1.

Intra-island genetic variability of foxes is indicated by several interrelated statistics (Fig. 2a). First, the APD among all foxes on each island is calculated from a pairwise difference matrix in which the proportion of restriction fragments that differ between any two individuals is given^{6,7}. For each island, a low APD indicates that individuals from that island share, on average, a greater proportion of their restriction fragments. Similarly, low standard errors indicate a less-variable distribution of pairwise difference values. Finally, the proportion of restriction fragments that are fixed among all foxes on each island potentially reflects the degree of homozygosity among them.

The intra-island APD ranges from 0.0% to 25.3% (Fig. 2a), compared with the APD of 70% to 90% typically found in outbred populations of vertebrates⁸⁻¹⁴. The populations of two large islands, Santa Catalina (SCa) and Santa Rosa (SRo), have the highest intra-island APDs (25.3% and 23.7%, respectively), whereas those of the two smallest islands, San Miguel (SMi) and San Nicolas (SNi), have the lowest (4.7% and 0.0%, respectively), indicating that differences in intra-island APD among populations correlate with population size, a prediction of population genetic theory¹⁵. But although the two remaining islands, Santa Cruz (SCr) and San Clemente (SCI), are large, their populations have relatively low APDs (8.5% and 10.0%, respectively). These lower-than-expected APDs could indicate that we

overestimated the population sizes of these islands, or that the population was recently reduced to an extremely small size owing to localized chance events, such as disease or drought.

The complete absence of variability in restriction-fragment profiles, as found among San Nicolas foxes, has not been previously reported for outbred populations. In fact, even within deliberately inbred strains of laboratory mice, individuals show slight variation in their restriction-fragment profiles¹¹. San Nicolas is the second-smallest Channel Island and the most remote, because it is separated from the other islands by 80 to 100 km (Fig. 1). Because the population on Santa Catalina was founded at about the same time as that on San Nicolas and retains much genetic variability (APD = 25.3%), it seems that small population size and a high degree of geographic isolation (rather than a more-recent origin) are the cause of differences in the variabilities of populations on these islands. Another consequence of small population size is that population bottlenecks can have dramatic effects on genetic variability^{16,17}. In 1974 and 1980, the San Nicolas population was estimated to be of <100 individuals and of <150 individuals, respectively^{18,19}. These decreases can be attributed to the presence of feral cats on the island, which have been abundant since 1970. We cannot, however, exclude earlier demographic contractions due to climatic change, disease outbreaks or other unknown factors.

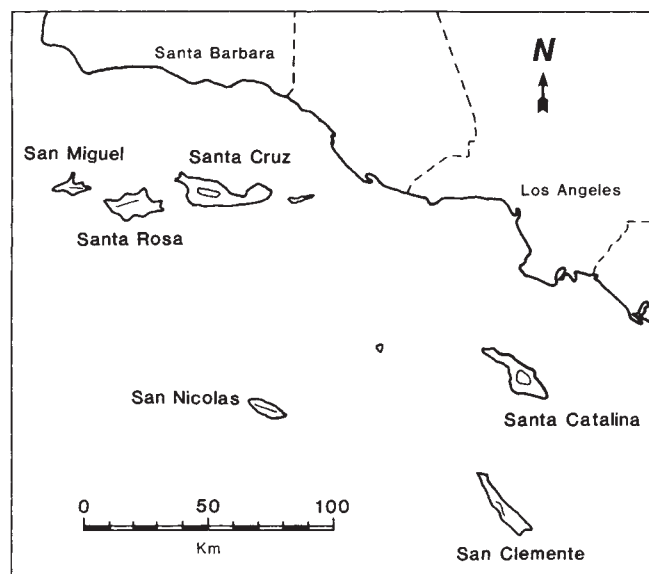


FIG. 1 Map of the California Channel Islands. Solid lines on islands indicate the transect along which fox samples were obtained. Area, population size, and time of establishment, respectively, of each island: San Miguel (37 km², 310, 9.5 kyr ago); San Nicolas (58 km², 490, 2.2 kyr ago); San Clemente (145 km², 1,230, 3.8 kyr ago); Santa Catalina (194 km², 1,650, 0.8–3.4 kyr ago); Santa Rosa (217 km², 1,850, 9.5 kyr ago); Santa Cruz (249 km², 2,120, 11.5 kyr ago). First record of an island fox, 16.5 kyr ago.

METHODS. On each island, 24–50 foxes were captured with cage traps. From each fox a blood sample was taken, and the fox was marked to avoid sampling recaptured individuals. Genomic DNA was isolated from blood samples of 10 to 16 individuals from each island, which were chosen out of the total sample to best represent the entire transect²³. Therefore, fox samples were derived from sampling 6–12 locations separated by 0.25–1 km along each transect. Population size is based on a density of 8.5 foxes per km², which is an average of density estimates from four islands for which accurate census data have been obtained (refs. 18, 19 and D. Garcelon, unpublished data). The range is 7.6 to 10.1 fox per km². The dates of establishment for the first island fox, and for foxes of San Nicolas, Santa Catalina and San Clemente, are based on radiometric dating of shell layers⁵. The first fox on Santa Catalina cannot be precisely dated because it occurs in an archaeological horizon bounded by layers dated at 0.8 and 3.4 kyr ago. The remains are much closer to the more recent horizon. The dates of origin for foxes on the northern islands, Santa Cruz, Santa Rosa and San Miguel, are based on the time of their final geographic separation from each other⁵.

§ To whom correspondence should be addressed.

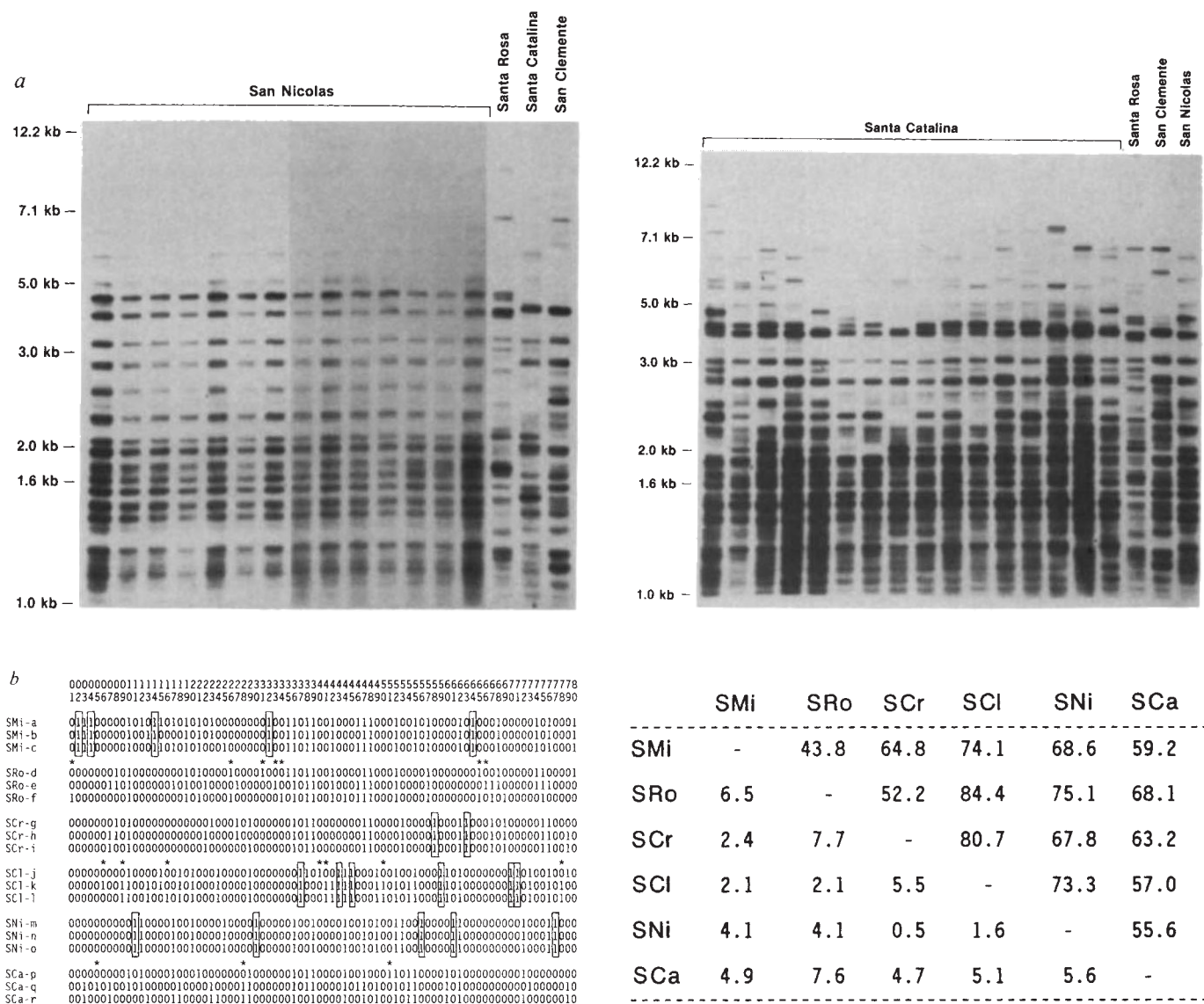


FIG. 2. *a*, Two examples of intra-island autoradiograms each of genomic DNA from 14–16 foxes. Autoradiogram of genomic DNA from San Nicolas (left) and Santa Catalina (right) foxes digested with *Hinf*I and probed with Jeffreys' 33.6 clone. Summary statistics derived from all six intra-island autoradiograms are as follows (sample size, *N*; APD (%) with range in parenthesis; s.e.m. with unbiased estimate in parenthesis; mean number of restriction fragments per individual, $\bar{n}$, with s.d. in parenthesis; proportion of fixed restriction fragments, *f*):

SMi: $N=10$; APD = 4.7(0.0–10.7); s.e.m. = 2.7(3.4); $\bar{n}=27.1(0.9)$; $f=0.92$.
 SNi: $N=14$; APD = 0.0; —; s.e.m. = 0.0(0.0); $\bar{n}=19.0(0.0)$; $f=1.00$.
 SCL: $N=16$; APD = 8.5(0.0–19.2); s.e.m. = 3.9(4.7); $\bar{n}=24.3(1.1)$; $f=0.82$.
 SCa: $N=16$; APD = 25.3(7.7–41.5); s.e.m. = 7.4(8.5); $\bar{n}=20.4(1.3)$; $f=0.39$.
 SRO: $N=16$; APD = 23.7(4.5–40.5); s.e.m. = 7.7(8.4); $\bar{n}=19.8(1.7)$; $f=0.35$.
 SCR: $N=11$; APD = 10.0(2.3–17.9); s.e.m. = 4.5(5.6); $\bar{n}=20.3(0.6)$; $f=0.79$.

b, Restriction fragment scores (top) and summary statistics (bottom) derived from analysis of an inter-island autoradiogram with three foxes from each of the Channel Islands. Top, presence-absence matrix of restriction fragments. Boxes surround restriction fragments uniquely shared by the three foxes on the same island. An asterisk indicates a unique fragment in one or two foxes on the same island. Bottom, APD matrix of foxes on different islands (above the diagonal), and the s.e.m. of ADP values (below the diagonal).

METHODS. DNA transfer and hybridization conditions were similar to those used previously⁷. In brief, 8 μ g DNA was digested with 50 U *Hinf*I in 100- μ l total reaction volume for 3 h at 37°C. After digestion, the sample was phenol-extracted, ethanol-precipitated, and resuspended in 17 μ l gel running buffer (TAE; 40 mM Tris, 20 mM sodium acetate, 1 mM EDTA at pH 7.2) for electrophoresis in 1% agarose gels with a 20-slot 1-mm-wide comb at 70 V

in 1 × TAE running buffer until all fragments <1 kilobase (kb) had run off the gel. The DNA was transferred onto Biotrace RP nylon membrane (Gelman Bioscience) and prehybridized in 0.5 M sodium phosphate at pH 7.2, 7% SDS, 1 mM EDTA and 1% BSA for 1–2 h at 65 °C. Hybridization was performed in the same buffer with the addition of 1 × 10⁶ c.p.m. ml⁻¹ ³²P-labelled 33.6 probe for 12–16 h at 65 °C, washed twice in 2 × SSC, 0.5% SDS, at room temperature for 10 min, then washed twice in 0.1 × SSC, 0.5% SDS, at 50 °C for 30 min. Several exposures varying from 1 to 10 days were made of each membrane to score accurately bands of differing intensity. All bands corresponding to the range of 1.0 to 12 kb were scored; those showing a similar relative molecular mass and intensity were considered to be identical. Only individuals on the same gel were compared. Relative molecular mass markers and duplicate samples from one individual were run on either side of the gel to control for distortion, which was minimal. As an experimental control, replicate gels were run and scored separately for foxes from San Nicolas, Santa Cruz, San Miguel, San Clemente, and for the inter-island analysis, at UCLA by N.L. and at the National Cancer Institute by D.A.G. The APD values from the two laboratories differed by <4% for the intra-island analysis, and the derived topologies for the inter-island analysis were the same in both cases, although branch lengths were slightly different. The difference value (*D*) was calculated as the number of fragments that differed between two individuals divided by the total number of fragments present in the two individuals. The APD is simply the average of all *D* values for each island multiplied by 100 (see refs. 6, 7). The inter-island APD is the mean of the *D* values between individuals in any two populations. Because of the non-independence of pairwise *D* values, an unbiased estimation of the s.e.m. may be defined as $2D(1-D)/(1+D)/\bar{n}(3+D)$, where $\bar{n}$ = mean number of restriction fragments per individual (M. Lynch, personal communication).

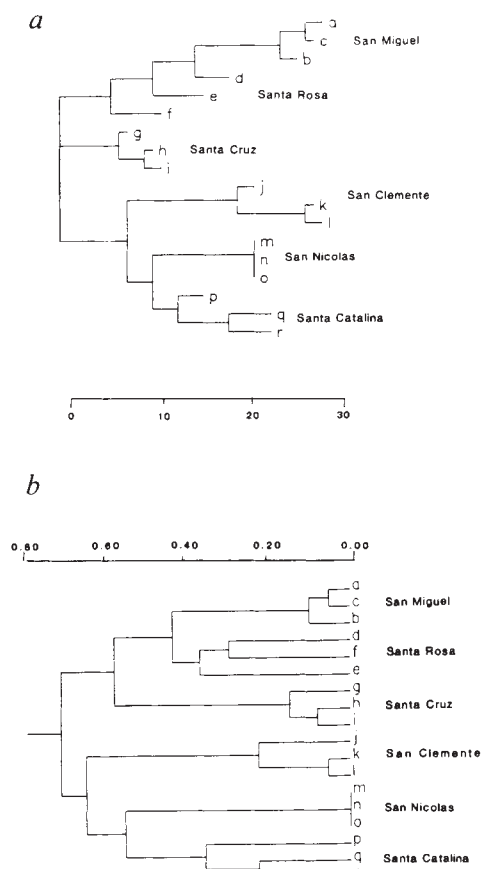


FIG. 3 *a*, Single most parsimonious tree of island foxes based on the presence-absence matrix in Fig. 2b. The tree was generated using the branch-and-bound option contained in Version 2.4 of the PAUP program of David Swofford (Illinois Natural History Survey, Champaign) and rooted at the midpoint of the longest path connecting any pair of taxa. Tree length = 135, consistency index = 0.593. Scale refers to the number of character-state (restriction-fragment) differences among taxa. *b*, UPGMA tree based on APD values above the diagonal of the matrix of Fig. 2b, bottom. Tree generated using the numerical taxonomy system (NTSYS) by James Rohlf (Applied Biostatistics Inc., Setauket, New York). Scale refers to the APD values among taxa.

Inter-island comparisons of restriction fragments indicate that the level of genetic variability between populations on different islands is higher than that within a population on a single island (Fig. 2). The APD between populations ranges from 43.8% to 84.4%, compared with 0.0% to 25.3% among foxes on a single island. San Miguel and Santa Rosa, the two islands least distance apart, are the most similar. These data indicate a considerable intra-island component to the evolution of restriction-fragment profiles. In the absence of gene flow, isolated small populations apparently become very distinct at hypervariable minisatellite loci.

Unique restriction fragments, found among all individuals on an island but not elsewhere, occur for fox populations of four islands (Fig. 2b). Fox populations on the other two islands, Santa Rosa and Santa Catalina, have the most-variable restriction-fragment profiles, so we might expect an absence or decreased number of fixed markers. Nevertheless, island-specific restriction fragments are present in each population although they are not shared by all individuals (Fig. 2b). Consequently, foxes of unknown origin can be assigned to a given island based on APD values or the presence of characteristic restriction fragments. We conclude that in small geographically isolated populations, DNA fingerprinting can be used to determine the population membership of unknown individuals.

The pattern of inter-island APD values and the presence of uniquely shared fragments among foxes on two or more islands (Fig. 2b), indicates that restriction-fragment data can be used to reconstruct phylogeny. Trees based on a maximum parsimony analysis of presence-absence of restriction fragments, and a phenetic analysis based on APD values, are nearly identical in their topology (Fig. 3). These trees group the foxes from the two closely situated northern islands, Santa Rosa and San Miguel. Santa Cruz foxes are not closely related to foxes on any other island. Foxes from the southern islands form a single monophyletic group in which foxes on Santa Catalina and San Nicolas are most closely allied. The groupings indicated by the trees in Fig. 3 are consistent with the archaeozoological and geological record: the fossil record indicates that foxes first arrived on the northern Channel Islands earlier than 16,500 years ago, having dispersed over water from the mainland to Santa Cruz⁵. At that time, the distance between the mainland and Santa Cruz was about 6.5 km and the three northern islands were joined. Santa Cruz separated from the other northern islands first, about 11,500 years ago, followed by the separation of Santa Rosa and San Miguel about 9,500 years ago (Fig. 1).

Of the southern islands, San Clemente has the oldest record of foxes beginning ~3,800 years ago, and San Nicolas has the second oldest record beginning 2,200 years ago⁵. The earliest fox on Santa Catalina is not precisely dated but comes from an archaeological site inhabited from 3,400 to 800 years ago. Taken together, the fossil record and the restriction-fragment trees indicate that San Clemente was the first southern Channel Island colonized by foxes, followed by San Nicolas and Santa Catalina. The southern populations were probably founded as a result of transportation by native Americans who had begun to colonize the islands by 10,000 years ago.

The difficulty in finding a population-level metric for comparing genetic variability of closely related populations and reconstructing their phylogenetic relationships, has been that most genetic markers, such as those revealed by protein electrophoresis, evolve at such slow rates that closely related populations are often too similar to discriminate. Because the mitochondrial genome evolves 5 to 10 times faster than a typical nuclear gene in mammals, mitochondrial DNA markers can be used effectively to reconstruct the phylogenetic history of populations^{20,21}. But mitochondrial DNA polymorphisms do not provide information about the extent of nuclear gene flow or variability. Our results indicate that restriction-fragment polymorphisms revealed with multilocus hypervariable minisatellite probes can be used as a quick and cheap way of estimating the relative genetic variability and reconstructing the phylogeny of small closely related, but isolated, populations. □

Received 15 November 1989; accepted 13 February 1990.

1. Lynch, M. *Molec. Biol. Evol.* **5**, 584-599 (1988).
2. Burke, T. *Trends Ecol. Evol.* **4**, 139-144 (1989).
3. Lewin, R. *Science* **243**, 1549-1551 (1989).
4. Hall, E. R. *Mammals of North America* Vol. 2 (Wiley, New York, 1981).
5. Collins, P. W. in *Recent Advances in California Islands Research*, (ed. Hochberg, F. G.) (Proc. Third Channel Island Symposium, Santa Barbara, California) (in the press).
6. Yuhki, N. & O'Brien, S. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 836-840 (1990).
7. Gilbert, D. A. et al. *Am. J. hum. Genet.* (in the press).
8. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76-79 (1985).
9. Burke, T. & Bruford, M. W. *Nature* **327**, 146-152 (1987).
10. Wetton, J. H., Carter, R. E., Parkin, D. T. & Walters, D. *Nature* **327**, 147-149 (1987).
11. Jeffreys, A. J. et al. *Nucleic Acids Res.* **15**, 2823-2836 (1987).
12. Jeffreys, A. J. & Morton, D. B. *Anim. Genet.* **18**, 1-15 (1987).
13. Georges, M. et al. *Cytogenet. Cell Genet.* **47**, 127-131 (1988).
14. Burke, T., Davis, N. B., Bruford, M. W. & Hatchwell, B. J. *Nature* **338**, 249-251 (1989).
15. Wright, S. *Evolution and the Genetics of Populations* Vol. 2 (University of Chicago Press, Chicago, 1969).
16. O'Brien, S. J. & Everman, J. F. *Trends Ecol. Evol.* **3**, 254-259 (1988).
17. Nei, M., Maruyama, T. & Chakraborty, R. *Evolution* **29**, 1-10 (1975).
18. Laughlin, L. in *The California Islands: Proceedings of a Multidisciplinary Symposium* (ed. Power, D. M.) 745-749 (Proc. Second Channel Island Symposium, Santa Barbara, California, 1980).
19. Kovach, S. D. & Dow, R. J. *Island Fox Research on San Nicolas Island* (Annual Report, Department of the Navy, Pacific Missile Test Center, Pt. Mugu, California, 1985).
20. Brown, W. M., George, M. & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1967-1971 (1979).
21. Avise, J. C. et al. *A. Rev. ecol. Syst.* **18**, 489-522 (1987).
22. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67-73 (1985).
23. Wayne, R. K. et al. *J. Heredity* **80**, 447-454 (1989).

ACKNOWLEDGEMENTS. We thank Alec Jeffreys for hypervariable minisatellite probes 33.15 and 33.6, and logistical support was provided by the United States Department of the Navy, Pt. Magu and North Island Naval Air Stations, the United States National Park Service, Channel Islands Region Office, the Nature Conservancy, the Santa Cruz Island Preserve, and the Santa Catalina Island Conservancy. For field assistance we thank L. Block, P. Collins, D. Garcelon, S. George, D. Girman, B. Gonzales, F. Hertel, S. Kovach, J. Malcolm, T. Martin, M. Murry, G. Romer, R. Rosomoff and L. Sadler. Financial support was provided by the Nature Conservancy and the Genetic Resources and Conservation Program of the University of California.

Gene for chronic proximal spinal muscular atrophies maps to chromosome 5q

J. Melki, S. Abdelhak, P. Sheth, M. F. Bachelot, P. Burlet, A. Marcadet*, J. Aicardi, A. Barois†, J. P. Carriere‡, M. Fardeau||, D. Fontan¶, G. Ponsot#, T. Billette§, C. Angelini, C. Barbosa††, G. Ferriere ‡‡, G. Lanzii|||, A. Ottolini|||, M. C. Babron¶¶, D. Cohen*, A. Hanauer##, F. Clerget-Darpoux¶¶, M. Lathrop*, A. Munnich & J. Frezal**

Unité de Recherches sur les Handicaps Génétiques de l'Enfant, INSERM U-12, et Département de Pédiatrie, Hôpital des Enfants Malades, 149 rue de Sèvres, 75743 Paris Cedex 15, France

* Centre d'Etudes des Polymorphismes Humains, 27 rue Juliette Dodu, 75010 Paris, France

† Hôpital Raymond Poincaré, 92380 Garches, France

‡ Hôpital de Purpan, 31052 Toulouse, France

|| Hôpital de la Salpêtrière, 75013 Paris, France

¶ Hôpital d'Enfants, 33000 Bordeaux, France

Hôpital Saint Vincent de Paul, 75014 Paris, France

§ Hôpital Trousseau, 75012 Paris, France

** Instituto di Clinica del Malattie Nervose e Mentali, 35128 Padova, Italy

†† Hospital de Crianças Maria Pia, 4000 Porto, Portugal

‡‡ Hôpital Saint-Luc, 1200 Bruxelles, Belgique

||| Università degli Studi di Pavia, 27100 Pavia, Italy

¶¶ INSERM U-155, Chateau de Longchamp, 75016 Paris, France

Institut de Chimie Biologique, 67085 Strasbourg, France

PROXIMAL spinal muscular atrophies represent the second most common fatal, autosomal recessive disorder after cystic fibrosis¹. The childhood form is classically subdivided into three groups: acute Werdnig-Hoffmann (type I), intermediate Werdnig-Hoffmann disease (type II) and Kugelberg-Welander disease (type III). These different clinical forms have previously been attributed to either genetic heterogeneity or variable expression of different mutations at the same locus². Research has been hindered because the underlying biochemical defect is unknown, and there are insufficient large pedigrees with the most common and severe form (type I) available for study. Therefore, we have undertaken a genetic linkage analysis of the chronic forms of the disease (types II and III) as an initial step towards the ultimate goal of characterizing the gene(s) responsible for all three types. We report here the assignment of the locus for the chronic forms to the long arm of chromosome 5 (5q12-q14), with the anonymous DNA marker D5S39, in 24 multiplex families of distinct ethnic origin. Furthermore, no evidence for genetic heterogeneity was found for types II and III in our study, suggesting that these two forms are allelic disorders.

Acute and chronic proximal spinal muscular atrophy (CP-SMA) are inherited in an autosomal recessive pattern in >90% of cases. The carrier frequency of autosomal recessive CP-SMA is ~1 in 90, whereas that of the acute form is 1 in 80¹. They are characterized by degeneration of anterior horn cells of the spinal cord, leading to progressive symmetrical limb and trunk paralysis, associated with muscular atrophy. Clinical heterogeneity of the different types of the disease in childhood has long been

recognized and various classification systems have been proposed²⁻⁴. Most of these systems divide childhood proximal SMA into types I, II and III, on the basis of a combination of age of onset, milestones of development, and age of survival.

We adopted the widely used and practical system of Dubowitz⁴ to distinguish families of type II and type III. Here, type II children are characterized by normal motor development up to the age of six months, with infants usually achieving the ability to sit, but unable thereafter to walk or stand unaided. Type III children have normal milestones in the first years of life and are able to walk unaided but show latent evidence of muscle weakness.

We collected a total of 24 multiplex families of European, Middle Eastern and African ancestry. Sixteen families were of type II and eight of type III (Fig. 1). Strict diagnostic criteria were followed: (1) proximal, symmetrical limb and trunk weakness; (2) muscle atrophy without facial or extraocular involvement; (3) no spasticity, hyperreflexia, sensory loss or mental retardation; (4) electromyographic studies showing denervation with normal nerve conduction velocity; (5) muscle biopsy consistent with denervation, with no evidence of storage material or other structural abnormalities.

We analysed 132 polymorphic DNA markers for linkage to CP-SMA. Markers were selected on the basis of their heterozygosity frequency of ≥30% and when possible, a genetic distance between them of ~20 centimorgans (cM). This study served to exclude the disease gene(s) from close proximity to these markers if a lod score ($\log_{10}$ odds ratio) of -2 or less was obtained⁵, usually at a recombination fraction of $\theta = \sim 0.10$.

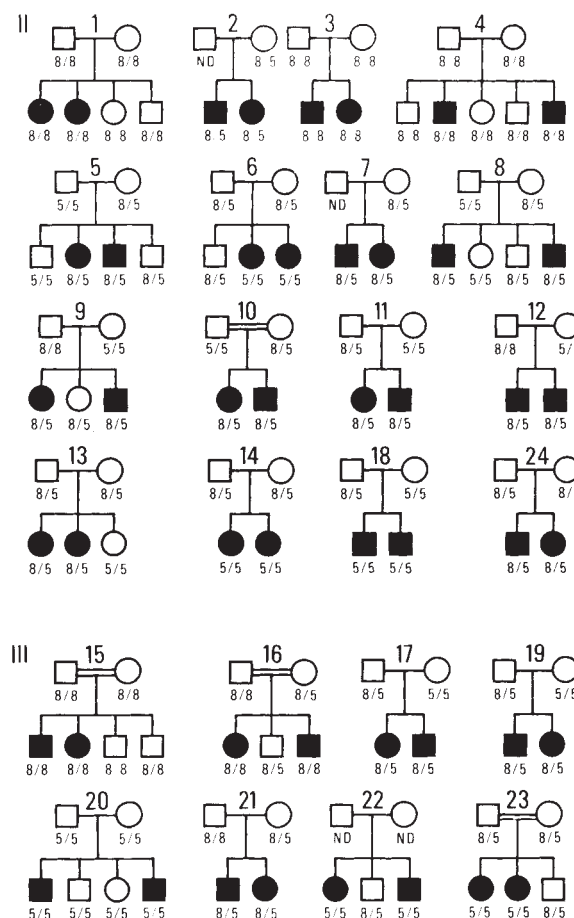


FIG. 1 CP-SMA pedigrees. Genotypes at the D5S39 locus are shown for each individual in affected families: type II (first four rows) and type III (the last two). RFLP detected with D5S39 was determined by hybridizing the DNA probe to Southern blots containing genomic DNA. ND, not done.

TABLE 1 Lod scores for linkage between D5S39 and CPSMA

Family	Number affected	Geographic origin	Recombination fraction (θ)			
			0	0.05	0.1	0.2
Type II						
2	2	Algeria	0.30	0.23	0.18	0.09
5	2	France	0.24	0.21	0.18	0.11
6	2	France	0.72	0.61	0.50	0.30
7	2	France	0.30	0.23	0.18	0.09
8	2	France	0.24	0.21	0.18	0.11
10	2*	France	0.30	0.25	0.21	0.13
11	2	France	0.30	0.25	0.21	0.13
13	2	France	0.42	0.32	0.22	0.08
14	2	Spain	0.60	0.51	0.42	0.26
18	2	France	0.30	0.25	0.21	0.13
24	2	France	0.30	0.21	0.14	0.05
Subtotal			4.06	3.35	2.69	1.53
Type III						
16	2*	Morocco	0.42	0.37	0.31	0.20
17	2	Belgium	0.30	0.25	0.21	0.13
19	2	Germany	0.30	0.25	0.21	0.13
21	2	France	0.30	0.25	0.21	0.13
22	2	Lebanon	0.48	0.41	0.34	0.22
23	2*	Tunisia	0.72	0.61	0.50	0.30
Subtotal			2.54	2.18	1.82	1.13
Total			6.60	5.53	4.51	2.66

The asterisk indicates the presence in the pedigree of inbred affecteds. Seven non-informative families are not shown (5 type II, 2 type III).

Assuming that these diseases result from a single-gene defect, we generated an exclusion map for >60% of the genome. Greater than 80% of chromosomes 3, 9, 10, 18 and 19, 40% of chromosomes 1, 2, 11, 15, 16, 17 and 21, and ~25% of chromosomes 5, 8, 12 and 14 were excluded (data available on request).

Linkage was demonstrated eventually between the mutation causing the disease and the anonymous probe D5S39 (plasmid p 105-153 Ra) located on the long arm of chromosome 5 (5q12-q14). Using the enzyme *Msp*1, this probe detects a restriction fragment length polymorphism (RFLP) of 5 kilobases (kb) and 8 kb, having frequencies of 60% and 40% respectively⁶. We assessed the potential genetic linkage of D5S39 to CP-SMA for all families using the computer program MLINK (version 4.9) of the LINKAGE package⁷. The results of this analysis are presented as lod scores in Table 1. The maximal lod score between CP-SMA and the locus D5S39 for combined sexes is +6.6 at a recombination fraction $\theta = 0.00$ (0-0.05 1-lod-unit confidence interval). No recombinants were observed between the CP-SMA locus and this probe. Subdivision of the families into types II and III, as previously described, resulted in lod scores of 4.06 and 2.54 respectively at $\theta = 0.00$.

These linkage data, derived from geographically isolated families and meeting strict diagnostic criteria, clearly establish that the defect causing types II and III SMA is located on the long arm of chromosome 5. In addition, our data show no evidence of heterogeneity between type II and type III as no recombinants were observed with D5S39 locus, suggesting that the two types are allelic disorders. Augmenting the number of informative meioses should establish if recombinations exist with this marker in each subgroup. At this point the use of flanking markers and multipoint analysis will become valuable.

It is possible and important to investigate the D5S39 locus in families with the type I form of the disease. Here, pedigrees are often very small because of the severe course of the disease, making primary linkage analysis difficult. Furthermore, continued insight into childhood SMA may now be gained from the Wobbler mutant mouse. This animal model displays a degenerative hereditary lower motor neuron disease similar to SMA⁸ and has previously been studied at the molecular level⁹. Given the existence of conserved homologous regions within

mammalian species, linkage of a marker in the Wobbler mouse chromosomal region analogous to human chromosome 5q12-q14 would prove extremely valuable.

Finally, for affected families, prenatal diagnosis will be shortly forthcoming, and the goal of gene characterization is now feasible. □

Received 5 February; accepted 15 March, 1990.

1. Pearn, J. *Lancet* **i**, 919-922 (1980).
2. Emery, A. E. H. *J. med. Genet.* **8**, 481-495 (1971).
3. Hausmanowa-Petrusewicz, I., Zaremba, J. A. & Borkowska, J. *J. med. Genet.* **22**, 350-353 (1985).
4. Dubowitz, V. *Brain* **87**, 707-718 (1964).
5. Morton, N. E. *Am. J. hum. Genet.* **7**, 277-318 (1955).
6. Leppert, M. *et al. Cytogenet. Cell Genet.* **46**, 649 (1987).
7. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3443-3446 (1984).
8. Duchon, L. W., Falconer, D. S. & Strich, S. J. *J. Physiol., Lond.* **183**, 53-55 (1966).
9. Murakami, T., Mastaglia, F. L., Mann, D. M. A. & Bradley, W. G. *Muscle & Nerve* **4**, 407-412 (1981).

ACKNOWLEDGEMENTS. We thank the Association Française contre les Myopathies (AFM), the families of patients and the physicians for their constant support; S. Lyonnet for advice, Y. Nakamura and R. White for providing polymorphic probes. Other probes including D5S39, were from the American Type Culture Collection (ATCC), and also Collaborative Research Incorporated. We thank L. Bonnet, D. Leleu, P. Pastureau, M. Masset, M. H. Herzberg and C. Massard for technical assistance. This study was also supported by the Ministère de la Recherche et de la Technologie.

Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification

A. H. Handside, E. H. Kontogianni, K. Hardy & R. M. L. Winston

Institute of Obstetrics and Gynaecology, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK

OVER 200 recessive X chromosome-linked diseases, typically affecting only hemizygous males, have been identified. In many of these, prenatal diagnosis is possible by chorion villus sampling (CVS) or amniocentesis, followed by cytogenetic, biochemical or molecular analysis of the cells recovered from the conceptus. In others, the only alternative is to determine the sex of the fetus. If the fetus is affected by the defect or is male, abortion can be offered. Diagnosis of genetic defects in preimplantation embryos would allow those unaffected to be identified and transferred to the uterus¹. Here we report the first established pregnancies using this procedure, in two couples known to be at risk of transmitting adrenoleukodystrophy and X-linked mental retardation. Two female embryos were transferred after *in vitro* fertilization (IVF), biopsy of a single cell at the six- to eight-cell stage, and sexing by DNA amplification of a Y chromosome-specific repeat sequence. Both women are confirmed as carrying normal female twins.

Five couples at risk of transmitting recessive X-linked diseases, including X-linked mental retardation, adrenoleukodystrophy, Lesch-Nyhan syndrome and Duchenne muscular dystrophy (DMD) were counselled about the possibility of having preimplantation embryos sexed after IVF and those identified as females replaced in the uterus. All of the women had previous terminations of affected fetuses but most had had difficulties in conceiving; each couple expressed a preference for this approach even though, in most cases, a specific diagnosis capable of distinguishing normal males and carrier females would be possible later in pregnancy. Over 6 months, 10 treatment cycles using superovulation after pituitary suppression² were performed (Table 1). Three couples had two treatments, and the other two couples, one and three treatments, respectively. The couples abstained from sexual intercourse during treatment cycles.

The number of oocytes collected and normally fertilized, 11.2 and 6.3 per treatment cycle, were consistent and similar to those we have previously reported². A high proportion of normally

TABLE 1 Number and sex of normally fertilized human embryos biopsied at the eight-cell stage

Patient/treatment	No. of oocytes	No. fertilized*	No. biopsied	Male	Female	Not sexed	No. transferred
A1	8	2	2	1	1	—	1
A2	9	6	6	3	3	—	2
B1	8	7	7	2	3	2	2
B2	8	4	3	1	2	—	1
B3†,‡	11	5	5	3	2	—	2
C1	12	8	5	2	3	—	2
C2	21	11	6	3	3	—	1
D1	14	10	7	4	2	1	2
E1	11	3	3	1	1	1	1
E2†	10	7	6	3	3	—	2
Total	112	63	50	23	23	4	17
(Range)	(8–21)	(2–11)	(2–7)	(1–4)	(1–3)		
Mean	11.2	6.3	5.0	2.3	2.3	0.4	1.7

* Normally fertilized with two pronuclei visible 16–18 h post insemination.

† Two eight-cell stage cells biopsied.

‡ Samples also reamplified with nested primers.

fertilized embryos (79%) developed to the six- to ten-cell stages on day 3 *in vitro* and were biopsied as early as possible. This allowed the biopsied embryos to be transferred later on the same day, after identification of the sex by DNA amplification (which took 6–8 h). For biopsy, a hole was drilled through the zona pellucida with acid Tyrodes (pH 2.4) applied locally from a fine micropipette. A larger micropipette was then introduced through this hole, and eight-cell stage cells aspirated for DNA analysis. In two treatment cycles, two cells were biopsied from each embryo for duplicate amplification (Table 1, B3, E2). Apart from the reduction in cellular mass, the development of human embryos to the blastocyst stage *in vitro* is unaffected by biopsy and removal of one or two cells at the eight-cell stage³. After uterine transfer, human cleavage stage embryos with reduced numbers of cells, for example after freezing and thawing⁴, can implant and develop normally.

Cell biopsies were examined for nuclei using interference contrast microscopy or vital labelling of DNA with a polynucleotide-specific fluorochrome and fluorescence microscopy. Labelling with the fluorochrome and brief exposure to ultraviolet light does not interfere with DNA amplification (E.H.K., K.H. & A.H.H., manuscript in preparation). In one case, this revealed that a cell biopsy was an anucleate cytoplasmic fragment and the corresponding embryo, therefore, could not be analysed (Table 1, E1). Despite lysis of some cell biopsies after removal, the nucleus remained intact and DNA amplification was unaffected.

Amplification of a short fragment of a Y-specific repeat sequence⁵ was consistently detected with dilutions of male DNA down to the equivalent amount in a single cell (2 pg), with no detectable background contamination (Fig. 1, male DNA). Amplification from single male cell biopsies was easily detectable and in some cases exceeded the levels achieved with a similar amount of male DNA (Fig. 1, biopsy 2). With other cells, no amplification was detectable and these were classified as female. Weak hybridization between the ends of the Y-specific primers results in amplification of a prominent 'primer-dimer' fragment (Fig. 1, bottom arrow) and provided a crude control for amplification failure. Even when background contamination increased, quantitative differences were sufficient for identification of the sex of the biopsied embryos. In a few cases, a regular ladder of amplified fragments, possibly arising from multimerization of the primers, prevented analysis (Table 1, B1, D1). Duplicate amplification from two cell biopsies with or without reamplification using nested primers (Table 1, B3, E2) did not significantly add to the discrimination of male and female embryos.

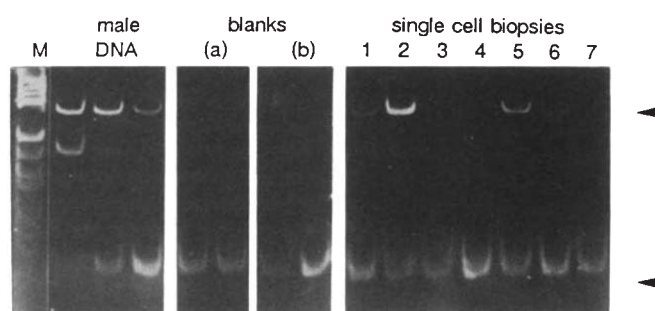


FIG. 1 Amplification of Y chromosome-specific DNA. Polymerase chain reaction (PCR) products using oligonucleotide primers for a 149 base pair (bp) fragment^{5,6} of a 3.4 kilobase (kb) sequence (DYZ1)⁸ repeated 500 to 8,000 times on the Y chromosome were analysed by polyacrylamide gel electrophoresis and stained with ethidium bromide. DNA size markers (M); serial dilutions of male DNA: 20 ng, 200 pg and 2 pg; wash drop buffer (a) and water (b) blanks; single cell biopsies from seven cleavage stage embryos (Table 1, D1). All of the samples were amplified and run together. Y-specific fragment (top arrow) (a smaller nonspecific fragment was amplified with 20 ng male DNA); 'primer-dimer' (bottom arrow). Among the single cell biopsies, the amplified fragment was clearly detectable with four cells (1, 2, 5, 6), in a fifth, several fragments of smaller sizes were visible which prevented analysis (4) and in the remaining two cells (3, 7) no amplification was detected.

METHODS. Sample preparation and PCR were as described previously⁶ with some modifications. Single cell biopsies were washed twice in PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.01% (w/v) gelatin) transferred to 10 µl water in 0.5 ml Eppendorf tubes, using a finely drawn pipette under a dissecting microscope to observe the cell entering the tube, and frozen to -20 °C. Samples were thawed, overlaid with 100 µl mineral oil and heat-denatured at 95 °C for 35 min. PCR mix (90 µl) was added to each sample (1 µg each oligonucleotide primer, 2.5U Taq polymerase and 200 µM each dNTP in PCR buffer), the DNA denatured for 4 min at 94 °C, followed by 40 cycles at 62–65 °C (90 s) and 94 °C (30 s), and a final extension at 72 °C (15 min) using an automated hot block (Perkin-Elmer Cetus). For reamplification, 2 µl original amplification mix was removed after 15 cycles and added to 100 µl fresh PCR mix containing nested primers (5'-ATTACACTACATTCCTTCCA-3'; 5'-AGTGAAATTGTATGCACTAGA-3') and amplified for an additional 40 cycles. Amplification product (10 µl) was run on a 12% polyacrylamide minigel, stained with ethidium bromide and examined on an ultraviolet transilluminator. White cells separated from whole blood samples from both partners were tested in each case for the presence and absence of the repeat normally located on the long arm of the Y chromosome⁸. In rare cases, this region is deleted in males⁹ and also it can be translocated to autosomes and carried by women¹⁰.

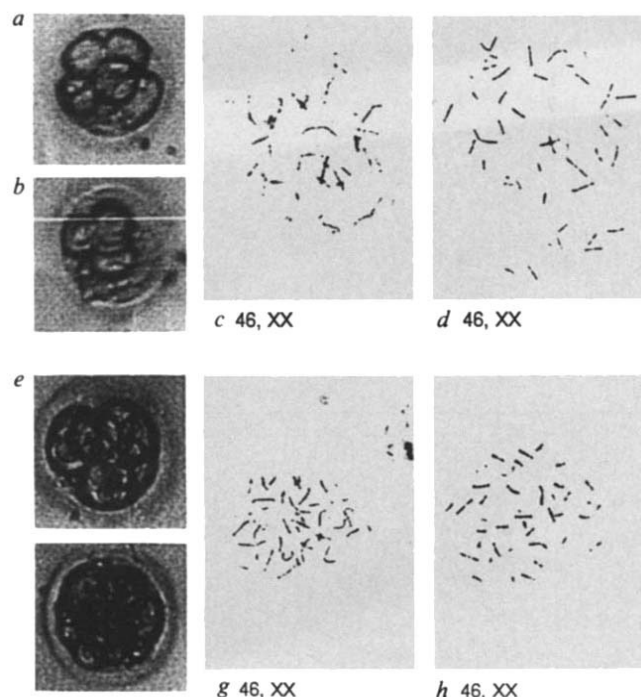


FIG. 2 The two pairs of biopsied embryos identified as female and transferred to the two women pregnant with twins. Patient/treatment cycle as follows. D1: a, seven-cell embryo (one cell removed and another damaged); b, seven-cell embryo (one cell removed and two others damaged). A2: e, six-cell embryo (one cell removed); f, eight-cell embryo (one cell removed). (Photographs of video image.) Metaphase spreads of banded chromosomes from cultured chorion villus cells recovered from each conceptus at ~10 weeks (D1: c, d; A2: g, h).

With the consent of the patients, the sex of some biopsied embryos (not selected for transfer) was later checked by amplifying the DNA of the embryo itself. With ten embryos (six male and four female), the predicted sex was confirmed. We have previously demonstrated the accuracy of sexing of normally fertilized embryos independently by *in situ* hybridization and fluorescent labelling of metaphase chromosomes⁶.

The sex ratio among the 46 biopsied embryos, which were classified, was exactly 50:50 but there was variation between batches of embryos. Although the probability of an embryo being female is only 0.5, in most treatment cycles, there were at least two female biopsied embryos for transfer. We consider this to be the maximum safe to transfer. CVS samples can be obtained from each twin with only a small risk of miscarriage whereas screening greater numbers of fetuses would be difficult.

Chorionic gonadotrophin (hCG) levels were assayed from 12 days after transfer; raised levels indicated implantation and early development in five cases. In two of these cases (Table 1, A1, B1), hCG levels were not maintained, indicating a 'biochemical' pregnancy. In a third case (Table 1, B3), serum hCG has continued to rise but clinical pregnancy is not yet confirmed. In each of the two remaining cases (Table 1, A2, D1), ultrasonography confirmed the presence of fetal sacs at 3 weeks and two fetal hearts at 4 weeks. These results suggest that pregnancy rates after transfer of biopsied six- to eight-cell embryos might equal those for intact embryos transferred to infertile patients. Within a 6 month period, pregnancies have been established in two (and possibly three) of the five women treated and after only one or two treatment cycles in each case.

CVS was performed in both sets of twins at 10 weeks and cytogenetic analysis of banded metaphase chromosomes confirmed that each of the fetuses has a normal female karyotype (Fig. 2). Both pregnancies are being carefully monitored by

ultrasonography and all fetuses are normal on detailed examination at 22 and 20 weeks. Both sets of twins appear on ultrasound to be dizygotic presumably arising from the implantation and development of both the transferred biopsied embryos. This will need to be confirmed at birth by DNA fingerprinting, but if both sets are dizygotic, the successful development of four biopsied embryos out of 17 transferred (24%) indicates that preimplantation diagnosis will be a viable option for many families carrying genetic defects. Unique sequences closely linked to the cystic fibrosis gene and an exon of the dystrophin gene frequently deleted in DMD have been amplified from single embryonic cells⁷, and the prospects for these families are promising. □

Received 28 March; accepted 30 March 1990.

1. Perleth, R. & McLaren, A. in *Clinical Obstetrics and Gynaecology* Vol. 1 (ed. Rodeck, C.) 747-764 (Bailliere Tindall, London, 1988).
2. Rutherford, A. J. *et al. Br. med. J.* **296**, 1765 (1988).
3. Hardy, K., Martin, K. L., Leese, H. J., Winston, R. M. L. & Handyside, A. H. *Hum. Reprod.* (in the press).
4. Trounson, A. & Mohr, L. *Nature* **305**, 707-709 (1983).
5. Kogan, S. C., Doherty, M. & Gitschier, J. N. *Engl. J. Med.* **317**, 985-990 (1987).
6. Handyside, A. H. *et al. Lancet* **i**, 347-349 (1989).
7. Coutelle, C. *et al. Br. med. J.* **299**, 22-24 (1989).
8. Cooke, H. J. *Nature* **262**, 182-186 (1976).
9. Bobrow, M., Pearson, P. L., Pike, M. C. & El-Alfi, O. S. *Cytogenetics* **10**, 190-198 (1971).
10. Cooke, H. J. & Noel, B. *Hum. Genet.* **67**, 222-224 (1979).

ACKNOWLEDGEMENTS. We thank Dr I. Soussis, Karin Dawson, Joe Conaghan and the IVF team at the Wolfson Family Clinic, Hammersmith Hospital; Professor Charles Rodeck for chorion villus sampling and Graham Davis for the cytogenetic analysis. This work was supported by Mr and Mrs G.-R. Flick, the Wolfson Family Charitable Trust and the Muscular Dystrophy Group of Great Britain and Northern Ireland, and was approved by the Research Ethics Committee of the Royal Postgraduate Medical School and the Interim Licensing Authority for Human In Vitro Fertilization and Embryology.

Arterial dilations in response to calcitonin gene-related peptide involve activation of K^+ channels

Mark T. Nelson, Yu Huang, Joseph E. Brayden, Jürgen Hescheler* & Nicholas B. Standen†

Department of Pharmacology and Vermont Center for Vascular Research, University of Vermont, Burlington, Vermont 05405, USA

* Institut Pharmakologie, Freie Universität Berlin, D-1000 West Berlin, FRG

† Department of Physiology, University of Leicester, Leicester LE1 7RH, UK

CALCITONIN gene-related peptide (CGRP) is a 37-amino-acid peptide produced by alternative processing of messenger RNA from the calcitonin gene^{1,2}. CGRP is one of the most potent vasodilators known³. It occurs in and is released from perivascular nerves^{1,4} and has been detected in the blood stream⁵⁻⁷, suggesting that it is important in the control of blood flow^{8,9}. The mechanism by which it dilates arteries is not known. Here, we report that arterial dilations in response to CGRP are partially reversed by blockers of the ATP-sensitive potassium channel (K_{ATP}), glibenclamide¹⁰⁻¹² and barium^{10,13}. We also show that CGRP hyperpolarizes arterial smooth muscle and that blockers of K_{ATP} channels reverse this hyperpolarization. Finally, we show that CGRP opens single K^+ channels in patches on single smooth muscle cells from the same arteries. We propose that activation of K_{ATP} channels underlies a substantial part of the relaxation produced by CGRP.

Rabbit mesenteric arteries contracted by noradrenaline (10 μ M) were relaxed by CGRP (Fig. 1), with 1.2 nM causing half relaxation. We tested the effects of potassium-channel blockers on CGRP-induced dilations. Glibenclamide (1 μ M), a sulphonylurea that inhibits K_{ATP} channels in mesenteric arteries¹⁰, cardiac muscle, and pancreatic β -cells, and which is specific for this channel type^{11,14}, reversed part of the dilation to CGRP (Fig. 1a). Glibenclamide (1 μ M), which had no effect

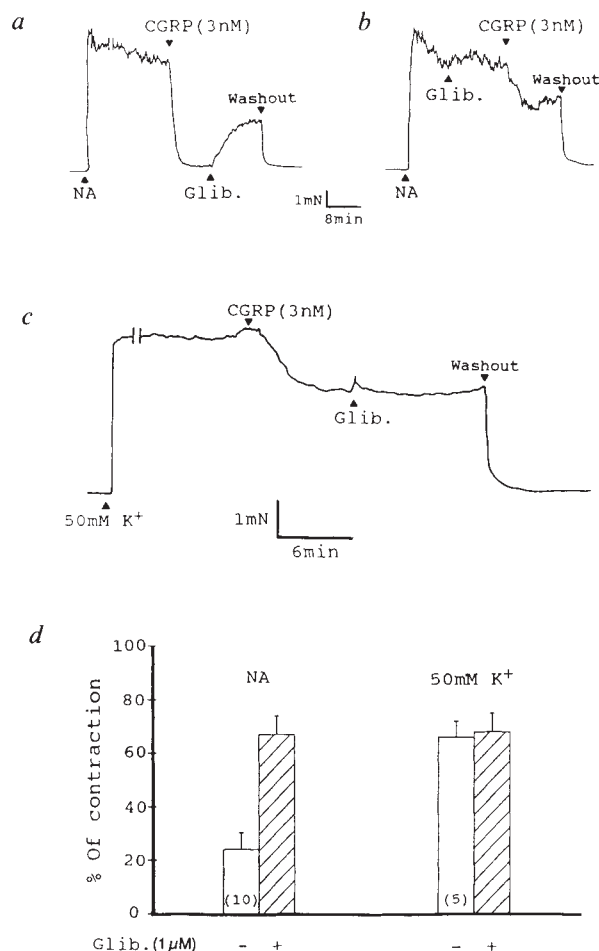
on the noradrenaline-induced contraction, reduced dilation in response to 3 nM CGRP by 57% (Fig. 1*b, d*). Glibenclamide at 1 μ M seemed to be a near maximally effective concentration as 10 μ M had little additional effect on the reversal of CGRP dilations ($n = 5$). Glibenclamide at 0.3 μ M caused $54\% \pm 6\%$ ($n = 5$) of its maximal reversal of 3 nM CGRP dilations. External barium blocks K_{ATP} channels at low concentrations¹³ and reversed 36% (mean of three experiments) of the CGRP-induced dilation at concentrations (50 μ M) that do not affect noradrenaline contractions¹⁰. In arteries contracted by raised potassium (50 mM), CGRP dilations were smaller than those in noradrenaline and unaffected by glibenclamide (Fig. 1*c, d*). These results suggest that the glibenclamide-sensitive dilation was due to a membrane hyperpolarization caused by K^+ channel activation.

Indeed, CGRP (5 nM) hyperpolarized smooth muscle cells in intact rabbit mesenteric arteries by 16.4 ± 1.5 mV (mean $\pm$ s.e.m.; $n = 10$), often taking the membrane potential close to the potassium equilibrium potential (about -85 mV) (Fig. 2*a*). CGRP (10 nM) has also been reported to hyperpolarize cat cerebral arteries by ~ 6 mV (ref. 15). CGRP-induced hyperpolarizations were reversed by glibenclamide (1 μ M, Fig. 2*a*). In the presence of CGRP, glibenclamide (1 μ M) depolarized by 19.8 ± 2.8 mV ($n = 5$). Glibenclamide (5 μ M) by itself caused a small depolarization (4.6 ± 1.5 mV; $n = 7$), suggesting that under these conditions K_{ATP} channels contribute to the membrane potential of smooth muscle cells. In the presence of glibenclamide (5 μ M), CGRP (20 nM) had no effect on membrane potential (Fig. 2*b*). In this experiment, after washout of glibenclamide, CGRP (10 nM) hyperpolarized this artery by 16 mV. Charybdotoxin (30 nM) (ref. 16) did not affect membrane potential or prevent the hyperpolarizing action of CGRP (Fig. 2*a*). In another experiment 30 nM charybdotoxin did not

reverse the 22 mV hyperpolarization to 5 nM CGRP (not shown). In concentrations of tetraethylammonium (TEA^+) (0.5–1 mM) that would block Ca^{2+} -activated K^+ channels^{17,18} by more than 70%, CGRP hyperpolarized these arteries by $18.8 \text{ mV} \pm 3.7 \text{ mV}$ ($n = 4$). These results suggest that CGRP hyperpolarizes arterial smooth muscle by opening K_{ATP} channels and not Ca^{2+} -activated K^+ channels.

To provide direct evidence that CGRP opens potassium channels, currents through single channels were measured with the patch clamp technique on single smooth muscle cells isolated enzymatically from rabbit mesenteric artery¹⁹. Cells were bathed in 120 mM K^+ solution to zero the membrane potential approximately, and the pipette solution contained 6 mM K^+ and 0.5 mM TEA^+ or 100 nM charybdotoxin to block Ca^{2+} -activated K^+ channels. In the experiment of Fig. 3*a*, occasional openings of a channel with an unitary current of 2.8 pA at 0 mV were seen under control conditions. CGRP (5 nM), when added to the bath solution, increased channel activity dramatically, so that up to four simultaneous openings were seen (Fig. 3*a, c*) while the unitary currents were unchanged. Channel activity can be expressed as the average number of channels open during the total duration of the measurement, which is equal to the product of the number of channels, N , in the patch and the open state probability of an individual channel, P_o . In this experiment, CGRP increased $N \cdot P_o$ 37.7-fold from 0.037 to 1.396, and assuming four channels in the patch, P_o increased from 0.009 to 0.349. CGRP (5 nM) increased P_o at 0 mV in seven experiments 28.8-fold from 0.017 ± 0.011 to 0.490 ± 0.115 (mean $\pm$ s.e.m.; $n = 7$) (the number of channels, N , in each patch was estimated from the maximum number of simultaneous openings observed in CGRP). As CGRP was applied only to the extra-patch membrane and not to the pipette solution, these results suggest that

FIG. 1 CGRP (rat, Sigma) relaxes mesenteric arteries from male rabbits. *a*, Glibenclamide (Glib., 1 μ M) reverses CGRP relaxation of an artery contracted by 10 μ M noradrenaline (NA). *b*, Glibenclamide (1 μ M) reduces the dilation to CGRP of an artery contracted by 10 μ M noradrenaline. Same artery as in *a*. After washing off the glibenclamide, CGRP (3 nM) dilated this artery to the same extent as in *a*. *c*, CGRP-induced dilation of an artery contracted by 50 mM potassium is insensitive to glibenclamide (1 μ M). Raising external K^+ would be expected to diminish the effects of any K^+ channel opener by substantially reducing the difference between the potassium equilibrium potential and membrane potential. *d*, Effect of 3 nM CGRP on noradrenaline (10 μ M) and high K^+ (50 mM) contractions in the presence and absence of glibenclamide (1 μ M). Data are normalized to the force level in the absence of CGRP. Hatched bars are the values in 1 μ M glibenclamide. The numbers in parentheses at the bottom of the open bars refer to the number of experiments on different arteries. The same number of experiments were done in the presence of glibenclamide (hatched bars). CGRP dilation of noradrenaline contractions was statistically different in the presence and in the absence of glibenclamide ($P < 0.01$) whereas CGRP dilation of high potassium contractions in the presence and absence of glibenclamide was not. CGRP at 1 nM without and with glibenclamide reduced force to 0.42 ± 0.07 ($n = 4$) and 0.63 ± 0.09 ($n = 4$) of control respectively. Isometric contractions of ring segments of mesenteric arteries were measured with a resistance vessel myograph²⁶. The solution bathing the arteries was in mM: 137 NaCl, 4.56 KCl, 0.44 KH_2PO_4 , 0.42 NaH_2PO_4 , 4.17 $NaHCO_3$, 11.1 glucose, 0.05 EGTA, 1 $MgCl_2$, 1.8 $CaCl_2$, 10 HEPES-NaOH, pH 7.4 at 37 °C. Glibenclamide (Sigma) was added to the bath solution from a concentrated stock (2 mM) dissolved in ethanol.



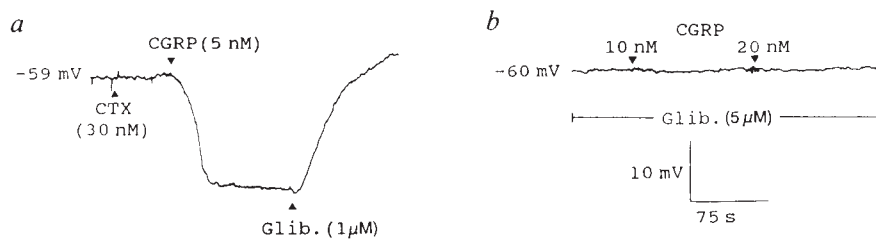


FIG. 2 CGRP hyperpolarizes rabbit mesenteric arteries. *a*, Glibenclamide (1 μ M) reverses CGRP-induced hyperpolarization. Charybdotoxin (CTX) did not affect membrane potential. *b*, Lack of effect on CGRP on membrane potential in the presence of glibenclamide (5 μ M). To measure membrane potentials, smooth muscle cells in the intact artery were impaled with a microelectrode under isometric conditions. Microelectrodes were filled with

0.5 M KCl and had tip resistances of 100–150 M Ω . For the membrane potential measurements, the arteries were not constricted with noradrenaline. Impalements are difficult to maintain in arteries when force is changing. The bathing solutions and temperature were the same as those used for the isometric force measurements.

it acts via an intracellular second messenger to increase channel P_o . The nature of the second messenger is not known, but may involve G proteins, as proposed for the activation of K_{ATP} channels by peptide hormones in insulin-secreting cells²⁰.

To characterize further the channel activated by CGRP, we measured the single channel current–voltage relation by recording channel activity at different potentials (Fig. 4*a*). The mean unitary current at 0 mV was $3.02 \text{ pA} \pm 0.11 \text{ pA}$ ($n=7$), with a slope conductance of about 20 pS between -50 mV and -30 mV, and the zero-current potential was negative to -60 mV (Fig. 4*b*). The single channel current–voltage relation was nonlinear, with the slope conductance increasing with membrane depolarization. The current–voltage relation was unchanged by patch excision into the inside-out configuration. Under these conditions, the cytoplasmic face was exposed to the 120 mM K^+ bath solution, giving a potassium equilibrium potential of -75 mV, whereas that for chloride ions is 0 mV. The channel

activated by CGRP was therefore potassium-selective. The reversal of the dilating and hyperpolarizing effects of CGRP by glibenclamide or barium suggests that CGRP activates the K_{ATP} channel that has recently been demonstrated in arterial smooth muscle¹⁰. Glibenclamide (30 μ M) applied to the bath reduced the activity ($N \cdot P_o$) of CGRP-activated channels in an on-cell patch from 3.12 to 0.37, without affecting the unitary current and in two other experiments on inside-out excised patches, glibenclamide (15 μ M) reduced the $N \cdot P_o$ of this K^+ channel to 0.37 and 0.36 of its control value respectively. The CGRP-activated channel clearly differs from the Ca^{2+} -activated K^+ channel of smooth muscle, which has a higher unitary conductance (at 0 mV, the unitary current was $5.40 \text{ pA} \pm 0.27 \text{ pA}$ ($n=6$) under identical conditions in this preparation). Consistent with the low affinity of external TEA^+ for K_{ATP} channels²¹, the unitary current of CGRP-activated channels was unchanged when TEA^+ (0.5 mM) was included in the pipette solution (Fig. 4*b*). TEA^+

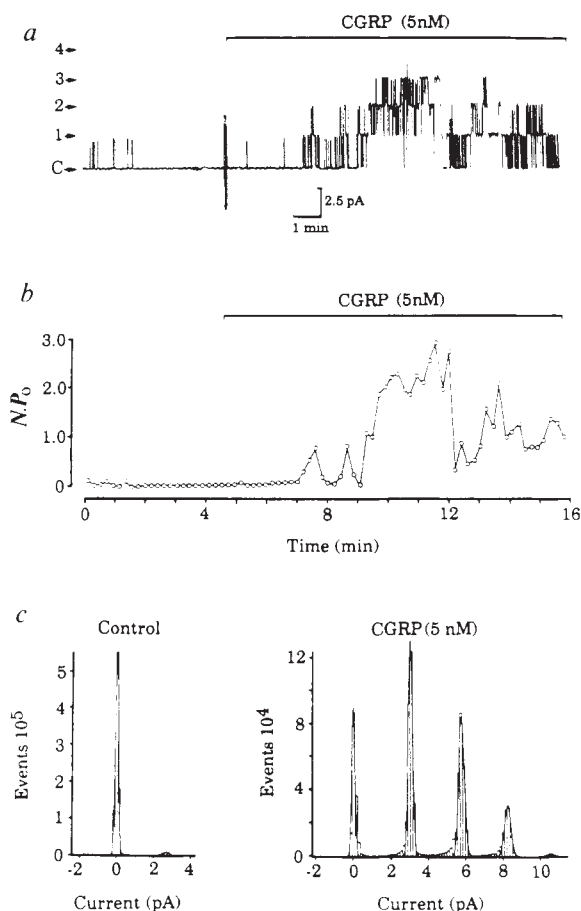


FIG. 3 CGRP opens single potassium channels. *a*, Continuous record at a potential of 0 mV of single potassium channels in an on-cell patch from a single smooth muscle cell of rabbit mesenteric artery. C, closed level. The numbers indicate the number of open levels. Final concentration of CGRP in bath solution was 5 nM. The displayed records were filtered at 0.5 kHz. The pipette solution in all CGRP experiments was 6 mM KCl, 114 mM NaCl, 1.8 mM $CaCl_2$, 10 mM HEPES-NaOH, pH 7.2. Charybdotoxin (100 nM) was also in the pipette. In two experiments, TEA (0.5 mM) was included in the pipette solution instead of charybdotoxin. This concentration of TEA blocks the Ca^{2+} -activated K^+ channel^{17,18} and had no effect on CGRP activation or single channel currents of K_{ATP} channels²¹. The bath solution was 104 mM KCl, 16 mM KOH, 1 mM EGTA, 1 mM $MgCl_2$, and 10 mM HEPES (pH 7.2). *b*, $N \cdot P_o$ from the experiment shown in *a*, calculated over 12.5 s intervals as $\sum_{j=1}^N (t_j)/T$, where t_j is the time spent in seconds with $j=1, 2, \dots, N$ channels open; N is the maximum number of channels seen (four in this case); and T is the duration of the measurement. *c*, Current amplitude histograms from the experiment in *a*. In the control, where most openings were very brief, unitary current amplitudes were measured only from events longer than 1 ms (three times the rise time of the 1 kHz filter²⁷). But the histogram illustrated was formed from the total recording in the control (450.9 s), as well as in CGRP (371.1 s), to illustrate the relative occupancy of the different levels. In controls and in CGRP, the probability of the closed, 1, 2, 3, or 4 levels being occupied were 0.964, 0.035, 0.001, 0.0, 0.0 and 0.189, 0.368, 0.307, 0.128, 0.008, respectively. Currents through single potassium channels were measured in patches on single smooth muscle cells from rabbit mesenteric artery with the patch clamp technique^{10,28}. Recordings were made with an Axopatch 1C amplifier and stored on video tape. The tapes were later replayed through an 8-pole Bessel filter and analysed with a Compaq 386s or 286 computer. Single channel experiments were performed at room temperature (18–22 $^{\circ}C$).

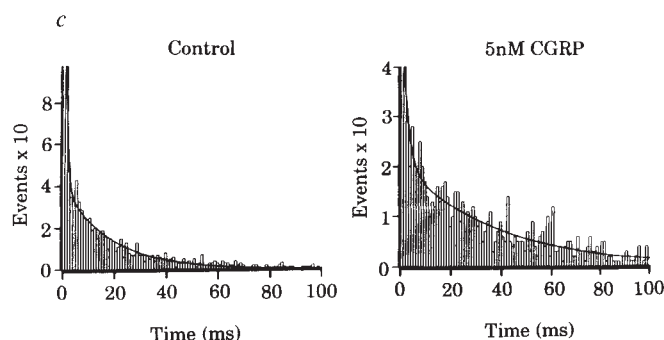
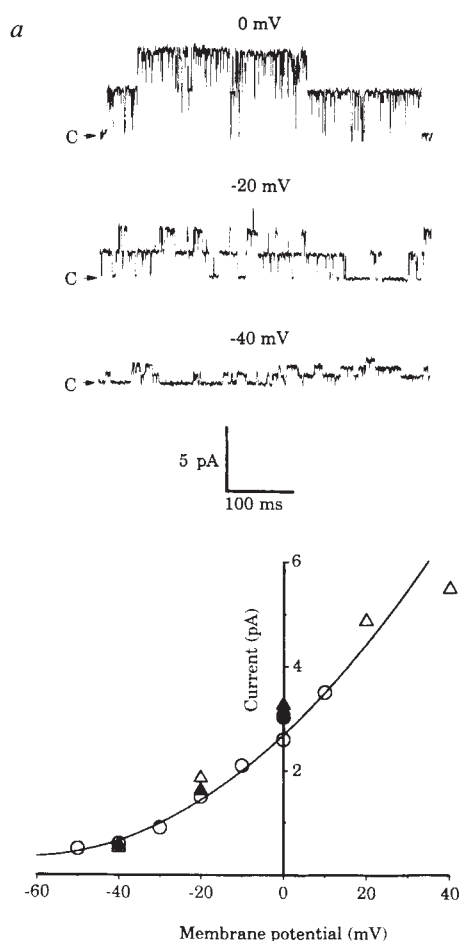


FIG. 4 Current-voltage relationship and open times of CGRP-activated potassium channels. *a*, Recordings at 0 mV, -20 mV, and -40 mV and filtered at 1 kHz. *b*, Single channel current-voltage relationship. The unitary currents were measured by fitting gaussian curves to the amplitude histogram using a least-squares algorithm. The pipette and bath solution compositions were the same as those used in the experiment shown in Fig. 3. The different symbols represent mean unitary currents of CGRP-activated channels from different experiments. The filled circle at 0 mV is the unitary current of channels activated by CGRP in the presence of 0.5 mM TEA⁺. The filled triangles are the unitary current values from channels activated by CGRP and inhibited by glibenclamide. *c*, Open time histograms from the experiment in Fig. 3, before and after 5 nM CGRP. Open time histograms were formed from idealized records using a half crossing approach. In each case the histogram was fitted by a least-squares method with two exponentials.

blocks single Ca²⁺-activated K⁺ channels in the same preparation with a dissociation constant of about 0.2 mM at 0 mV (ref. 18).

The open times of these potassium channels at 0 mV could be fitted by two exponentials, with time constants in control and CGRP of 0.9 ms and 17.9 ms, and 2.2 ms and 37.8 ms, respectively (Fig. 4c). This small increase in these two open time constants clearly cannot account for 38-fold CGRP-induced elevation of N_{Po} . K_{ATP} channels in inside-out excised patches from mesenteric artery cells in the absence of intracellular ATP and in the presence of 1 mM ATP and 1 μ M cromakalim also exhibit two open states, with similar time constants (~1 ms, ~15 ms). Two open states have also been reported for K_{ATP} channels in skeletal²² and cardiac muscle²³.

CGRP-containing nerves are part of a class of non-adrenergic,

non-cholinergic vasodilator nerves²⁴ in resistance vessels that may have an important role in regulation of blood pressure⁹. We demonstrate that CGRP hyperpolarizes arterial smooth muscle by activating K⁺ channels. This hyperpolarization seems to be responsible for part of the observed vasorelaxation. The mechanism of the membrane hyperpolarization-independent component of dilation remains to be determined and may involve calcium channel inhibition, shift in the intracellular Ca²⁺-force relationship, or increased Ca²⁺ extrusion or sequestration. There is evidence that cromakalim and pinacidil, members of a new class of pharmacological vasodilators^{14,25}, act by opening K_{ATP} channels¹⁰ as do vasoactive intestinal peptide and endothelium-derived hyperpolarizing factor¹⁰. We therefore propose that the K_{ATP} channel of smooth muscle may be a common target for a number of vasorelaxing agents, and thus an important contributor to vasodilation. □

Received 3 October 1989; accepted 19 February 1990.

- Rosenfeld, M. G. *et al.* *Nature* **304**, 129-135 (1983).
- Emeson, R. B., Hedgiran, F., Yeakley, J. M., Guise, J. W. & Rosenfeld, M. G. *Nature* **341**, 76-80 (1989).
- Brain, S. D., Williams, T. J., Tippins, J. R., Morris, H. R. & MacIntyre, I. *Nature* **313**, 54-56 (1985).
- Mulderry, P. K. *et al.* *Neuroscience* **14**, 947-954 (1985).
- Morris, H. R. *et al.* *Nature* **308**, 746-748 (1984).
- Edbrooke, M. R. *et al.* *EMBO J.* **4**, 715-724 (1985).
- McEwan, J. R. *et al.* *Circulation* **77**, 1072-1080 (1988).
- Brain, S. D. & Williams, T. W. *Nature* **335**, 73-75 (1988).
- Kawasaki, H., Tasaki, K., Saito, A. & Goto, K. *Nature* **335**, 164-167 (1988).
- Standen, N. B. *et al.* *Science* **245**, 177-180 (1989).
- Ashcroft, F. M. A. *Rev. Neurosci.* **11**, 97-118 (1988).
- Stanfield, P. R. *Trends Neurosci.* **10**, 335-339 (1987).
- Quayle, J. M., Standen, N. B. & Stanfield, P. R. *J. Physiol. Lond.* **405**, 677-697 (1988).
- Quast, U. & Cook, N. S. *Trends Pharmac. Sci.* **10**, 431-435 (1989).
- Saito, A., Masaki, T., Uchiyama, Y., Lee, T. J. F. & Goto, K. *J. Pharmac. exp. Ther.* **248**, 455-462 (1989).
- Miller, C., Moczydowski, E., Latorre, R. & Phillips, M. *Nature* **313**, 316-318 (1985).
- Villarroel, A., Alvarez, O., Oberhauser, A. & Latorre, R. *Pflügers Arch.* **413**, 118-126 (1988).

- Huang, Y., Langton, P. D., Hescheler, J. K., Standen, N. B. & Nelson, M. T. *Biophysical J.* **57**, 117a (1990).
- Nelson, M. T., Standen, N. B., Brayden, J. E. & Worley, J. F. *Nature* **336**, 382-385 (1988).
- Dunne, M. J., Bullett, M. J., Li, G. D., Wollheim, C. B. & Petersen, O. H. *EMBO J.* **8**, 413-420 (1989).
- Davies, N. W., Spruce, A. E., Standen, N. B. & Stanfield, P. R. *J. Physiol. Lond.* **413**, 31-48 (1989).
- Spruce, A. E., Standen, N. B. & Stanfield, P. R. *Nature* **316**, 736-738 (1985).
- Kakei, M., Noma, A. & Shibasaki, T. *J. Physiol. Lond.* **363**, 441-462 (1985).
- Bevan, J. A. & Brayden, J. E. *Circ. Res.* **60**, 309-326 (1985).
- Winquist, R. J. *et al.* *J. Pharmac. exp. Ther.* **248**, 149-156 (1989).
- Mulvany, M. & Halpern, W. *Circ. Res.* **41**, 19-26 (1977).
- Colquhoun, D. in *Microelectrode Techniques, The Plymouth Workshop Handbook* (eds Standen, N. B., Gray, P. T. A. & Whitaker, M. J.) 83-104 (Company of Biologists, Cambridge, 1987).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 85-100 (1981).

ACKNOWLEDGEMENTS. We thank Dr E. Moczydowski for the gift of purified charybdotoxin and Drs J. Patlak and J. Quayle for discussion and comments on the manuscript. This work was supported by grants from the NSF, the American Heart Association, NIH, MRC, Nuffield Foundation, and a NATO collaborative grant. N.B.S. is a Wellcome Trust Research Leave Fellow. M.T.N. is an Established Investigator of the American Heart Association.

TABLE 1 Sequence analysis

(a) Percentage identities between sequence pairs

	US27	US28	RHOD	5HTA	ADR	MAR	SKR	RDC7	MAS	CAR	STE3
US27			24	23	27	25	27	25	19	11	11
US28	40		27	27	26	29	31	24	18	17	11
UL33	24	24	19	22	20	23	20	18	17	16	10

(b) Statistical significance of HCMV and GCR sequence alignment

	Motif	Location	Class	Cutoff	Start	Range	Probability
1	N,,,V	TM I	6	1.0			0.106E+00
2	L,,,D	TM II	6	1.0	15	20	0.125E-00
3	C	Ex	1	1.0	25	15	0.290E-01
4	DE, R, Y	Cy	6	3.0	25	0	0.192E-03
5	W	TM IV	1	1.0	22	10	0.130E-01
6	P	Tm IV	1	1.0	10	1	0.521E-01
7	C	Ex	1	1.0	10	30	0.290E-01
8	P	TM V	1	1.0	18	15	0.521E-01
9	Y	TM V	1	1.0	9	1	0.340E-01
10	F,,,,,P	TM VI	6	2.0	20	160	0.188E-02
11	N,P,,,Y	TM VII	6	3.0	30	30	0.763E-04

In the comparison of (a) (see below), HCMV, US27 and US28 (40% identity) are more closely related than any other pair except for the β_2 -adrenergic (ADR) and 5-hydroxytryptamine (5HTA) receptors (47%). In comparisons with non-HCMV sequences, US27 and US28 share highest identities with the cellular hormone and neurotransmitter receptors, in particular the receptors for substance K (SKR) and muscarinic acetylcholine (MAR), and β_2 -adrenergic receptors. UL33 is most similar to US 27 and US28, but also highly similar to MAR and 5HTA. Identities were counted in the following aligned sequence blocks shown in Fig. 2: positions 53-71; 87-108; 123-149; 170-189; 230-242; 244-252; 300-322; and 342-365. These 157 positions mostly represent the bulk of the transmembrane domains and are unpadding except at point 185. Identities were counted using the MULTALIGN program in pairwise mode⁴¹. For further abbreviations, see Fig. 2 legend. In b, a set of motifs including most highly conserved residues was derived from the multiple alignment in Fig. 2. This pattern of motifs was designed to allow the identification of the first nine sequences in Fig. 2, with the specific aims of estimating the probability of the pattern occurring by chance and of determining if non-HCMV and non-GCR sequences could be found in the database using the pattern. As the pattern is minimalistic, the probability calculation largely excludes any contribution due to compositional (hydrophobic) bias, and minimizes nonrandom oligopeptide biases⁴². Pattern probability (product of the motif probabilities)=0.3693E-21, where E=exponent. Expected number of matches per 1,000 residues=0.1521E-07. Searches against the PIR (1,766,843 residues), PIRNEW (1,639,179 residues), and SWISSPROT (3,265,967 residues) databases identified known and putative G protein-coupled receptors and the HCMV US27 and US28 sequences (previously deposited in the databases). No other matches were found. The configuration of the program options used is described below. Two classes of motifs were used; 1 (exact match) and 6 (membership of a set). In the latter definition, residue positions are separated by commas, with alternative amino acids at a given position not being separated. The cutoff score represents the minimum match (in number of amino acids) that must be found. The search window of each motif is defined relative to the previous motif (except for motif 1). The start of the window is shown in the column 'Start', with numerals representing the distance downstream of the first position in the previous motif. The range specifies the size of the window. Single-letter amino-acid code used for motifs. Abbreviations: TM, transmembrane; Cy, cytoplasmic; Ex, extracytoplasmic.

in HCMV remains enigmatic, such gene families are transcriptionally active and some express protein¹⁴⁻¹⁹. All cellular gene homologues that have been investigated, at least in the human herpesviruses, seem to be expressed.

Further evidence for expression of UL33, US27 and US28 is that the first ATG codon in each reading frame conforms to the Kozak consensus²⁰. But promoter elements cannot be confidently predicted from the sequence data. Possible TATA boxes exist upstream of each frame, but deviate considerably from the consensus (TATTCGC, CATAAAA and TATTTT, occurring 199, 163 and 175 base pairs (bp) upstream of the first ATG of the UL33, US27 and US28 frames, respectively). A potential polyadenylation signal exists downstream of US28 (ref. 21), but not downstream of the other frames. This is common for other HCMV genes, some of which may be transcribed as 3'-coterminal families²¹.

It remains to be determined if the HCMV-GCR homology is reflected at the level of function. But residues that are functionally and structurally important in some cellular GCRs are also conserved in the viral sequences. The N-terminal extracellular domains of US27 and US28 contain potential glycosylation sites (four and one respectively), similar to most known GCRs (Fig. 2). By contrast, the putative UL33 protein, like the canine RDC7 and RDC8²² and yeast STE3 sequences, has a short N-terminal extracellular domain lacking potential glycosylation sites. Although glycosylation does not seem to play a direct part in receptor function, it may affect receptor localization²³.

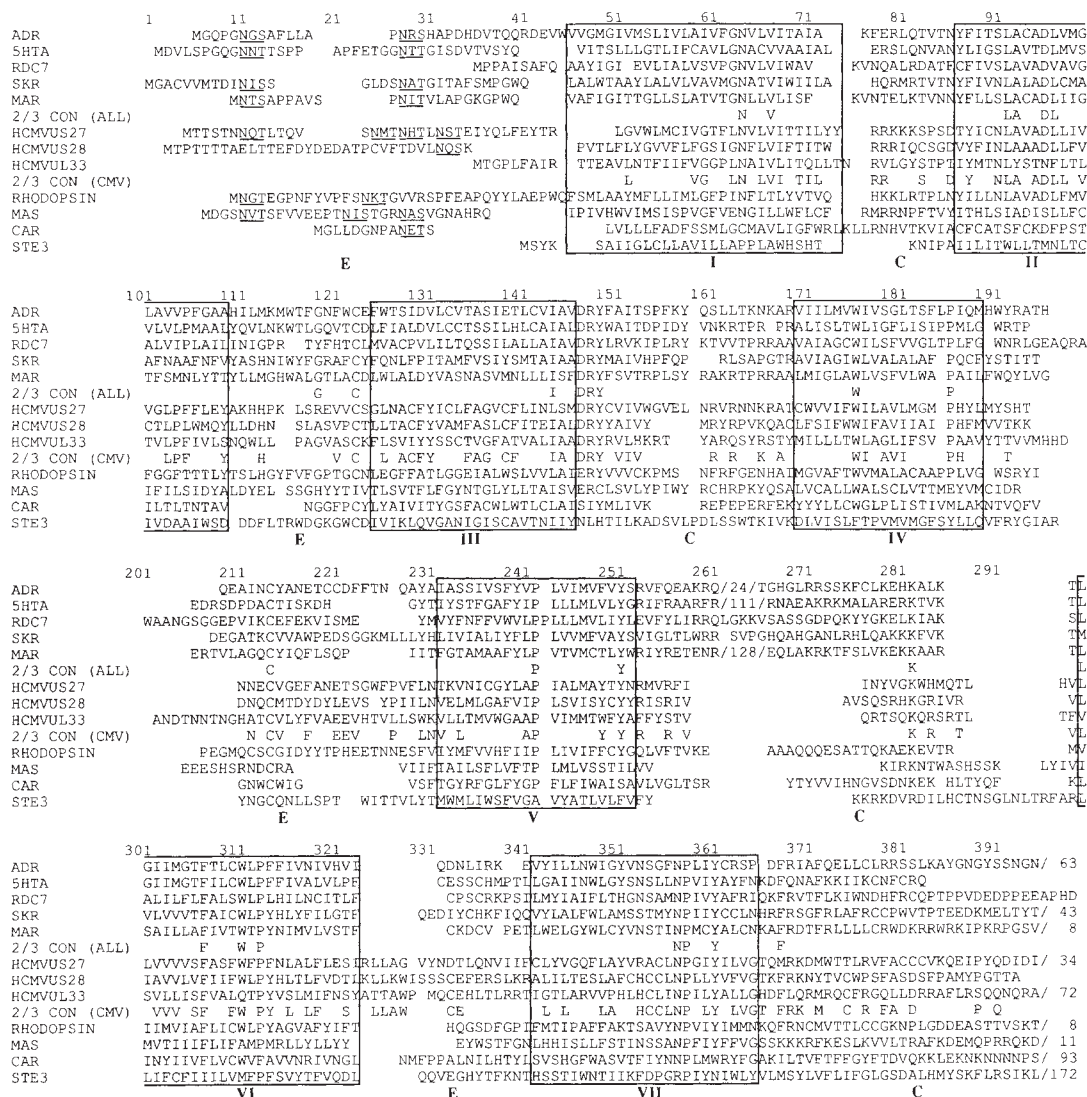
The second and third extracellular domains of GCRs contain two highly conserved cysteine residues (positions 123 and 214

of Fig. 2). Of the ten cysteine residues in bovine rhodopsin, only the two cysteines conserved with other GCRs are essential to the structure²⁴. The same two cysteines are also conserved in the HCMV sequences and equivalent residues are important in β -adrenergic receptor function²⁵.

The transmembrane domains feature conserved proline residues (in helices IV-VII; Fig. 2), also found in bacteriorhodopsin and ion channels. These residues may help constitute a pore or pocket, allowing the passage of ions or the binding of ligands, respectively²⁶. Proline 358 (numbering as in Fig. 2 numbering) of the β -adrenergic receptor, which is highly conserved in GCRs and the HCMV sequences, has been shown by mutagenesis to be structurally important²⁷. Ligand binding in adrenergic receptors is largely a function of the transmembrane helices^{28,29}. It is unknown if the HCMV sequences bind ligands, but they lack an aspartate residue in helix III which characterizes the cationic amine receptor subclass³⁰ (position 130 of Fig. 2). Aspartate 96, however, which is important for ligand-binding in the β -adrenergic receptor²⁷, is conserved in US27 and US28; UL33 has an asparagine at position 96, which could be regarded as a conservative change, on the basis of mutation probabilities.

G-protein interactions have been localized predominantly to the third cytoplasmic and C-terminal domains. For example, the substitution of lysine 248 of rhodopsin (position 282 of Fig. 2) with leucine abolishes its ability to activate transducin³¹. This lysine residue is located in a basic region, near the end of the third cytoplasmic domain, which is well conserved in the cellular GCRs. Although the alignment shown in Fig. 2 is relatively uncertain in this region, all three HCMV sequences possess a lysine residue, which, in US27 and US28, is found in relatively

FIG. 2 Alignment of the three HCMV-GCR homologues with nine known or putative G protein-coupled receptors. (Single-letter amino-acid code). The alignment shows only representative sequences of different branches of the GCR family. Approximate positions of the transmembrane domains (I-VII) are indicated by shading. Hydrophilic domains are labelled as C (cytoplasmic) or E (extracellular)^{47,48}. Potential N-linked glycosylation sites are underlined in the N-terminal cytoplasmic domain. Consensus lines show residues that are conserved in at least 8 of the 12 sequences (ALL) or 2 of the 3 HCMV sequences (CMV), only where sequences are unpadding. Despite different evolutionary histories, conserved positions in both sets of sequences are superimposed. Residues have been removed from the alignment in the third cytoplasmic domain of 5HTA, ADR and MAR, and the alignment amputated at the C-terminal end. In each case the number of residues omitted is shown. Abbreviations and sources: 5HTA (human 5-hydroxytryptamine_{1A} receptor)⁴⁹; ADR (human β_2 -adrenergic receptor)⁵⁰; RDC7 (canine putative GCR)²²; SKR (bovine substance-K receptor)⁵¹; MAR (porcine muscarinic acetylcholine receptor)⁵²; rhodopsin (bovine)⁵³; MAS (human MAS oncoprotein, angiotensin receptor)^{54,55}; CAR (*Dictyostelium discoideum* cAMP receptor)⁵⁶; STE3 (yeast *STE3* a factor pheromone receptor)⁵⁷. The cAMP and yeast pheromone receptor sequences are included to emphasise the extraordinary diversity of the receptor type,



basic surroundings. It is of interest that the third cytoplasmic domain in the HCMV sequences is shorter than that of the mammalian hormone receptors. An exception is the *mas* oncogene which clusters with the yeast, *Dictyostelium* and HCMV sequences in this respect.

In contrast to the third cytoplasmic domain, the HCMV sequences vary significantly in length in the C-terminal domain (Fig. 2). In cellular GCRs the C-terminus is phosphorylated by serine-threonine kinases, including specific GCR kinases, which results in the rapid desensitization of receptors³². The HCMV sequences also contain potential phosphorylation sites in their C-terminal domains, and they may therefore be amenable to control by cellular kinases (Fig. 3).

It seems likely that in acquiring genes for these receptor-like molecules, HCMV has evolved a way of using complex cellular-signalling mechanisms with a minimum of genetic investment. The variation between the genes indicates that, singly or in combination, they provide an adaptable response in the host environment. These proteins could regulate HCMV latency or reactivation, about which little is known. The HCMV major immediate early gene enhancer, which is involved in initiating

the lytic replication cycle, contains cyclic AMP response elements^{33,34}; some GCRs trigger the raising or lowering of cAMP levels in the cell. During viral replication in cultured fibroblasts, protein synthesis is up regulated³⁵; such a response could also be augmented by GCR signalling.

Perhaps the most interesting possibility arises from the association of cellular GCRs with roles in cell proliferation and control of development. Expression of the 5-hydroxytryptamine 1c

RHODOPSIN	G D D E A S T P V S K E E S Q V A P A
HCMVUS27	R H Y D R K N A P M E G E E E F L L
HCMVUS28	V C W P K A A D S F P A M Y P G T A
HCMVUL33	D S M K I P H R L Q Q S H N L S G V

FIG. 3 Potential phosphorylation sites in the C termini of the HCMV GCR homologues. Serine and threonine residues in the C-terminal 20 amino acids of the HCMV and bovine rhodopsin sequences are circled. Neighbouring charged residues are highlighted. Phosphorylation of at least some of these serines and threonines can occur in rhodopsin, which can also be phosphorylated in the third cytoplasmic domain⁴⁷.

receptor in NIH 3T3 fibroblasts results in cell transformation³⁶. Transformation is dependent on activation of the receptor by ligand, which is also required to maintain the transformed phenotype. GCRs could also have normal roles in development and cell proliferation⁴. Against this background, the homology of the three HCMV sequences to the *mas* oncogene (Fig. 2) and the possible association with G proteins, which can also be oncogenic³⁷, is noteworthy. HCMV has been implicated in malignancy, but no association has been proven³⁸.

The possibility that the viral proteins bind small molecules makes them potential targets for antiviral drugs^{39,40}. The prospect of specifically blocking the HCMV homologues is raised by their significant sequence differences to known cellular GCRs, and it is conceivable that suitable drugs already exist in the pharmacopoeia. Apart from complementing conventional antiviral chemotherapy, the availability of a disparate drug target could facilitate the control of resistant mutants. These possibilities are contingent on the demonstration that the HCMV proteins are functionally homologous to GCRs.

Note added in proof: The EMBL database accession number for the HCMV strain AD169 DNA sequence is X17403. □

Epstein-Barr virus latent membrane protein inhibits human epithelial cell differentiation

Christopher W. Dawson, Alan B. Rickinson
& Lawrence S. Young*

Cancer Research Campaign Laboratories, Department of Cancer Studies,
University of Birmingham Medical School, Birmingham B15 2TJ, UK

EPSTEIN-BARR virus (EBV), a human herpesvirus, is strongly linked with two relatively rare forms of B-cell lymphoma and with a much more prevalent epithelial malignancy, undifferentiated nasopharyngeal carcinoma (NPC)¹. The availability of suitable culture systems has allowed detailed analysis of EBV-induced growth transformation in B lymphocytes, but little is known about the virus-epithelial cell interaction or about the possible effector role of viral proteins in the pathogenesis of NPC. Here we describe an experimental system to monitor the effects of introduced viral or cellular genes upon human epithelial cell growth and differentiation. We transfected a human epithelial cell line, which retains several features of normal keratinocyte behaviour *in vitro*, with the EBV gene encoding latent membrane protein (LMP), one of only two viral proteins known to be expressed in NPC cells *in vivo*^{2,3}. LMP expression was accompanied by changes in the epithelial cell surface phenotype, mimicking surface changes observed in NPC cells, and by severe impairment of the cellular response to differentiation signals. The ability of LMP to inhibit terminal differentiation indicates a mechanism whereby EBV infection of squamous epithelium could contribute to the multi-step pathogenesis of NPC.

As active EBV infections have been difficult to establish in cultured human epithelium by direct exposure to virus preparations^{4,5}, we sought to develop an alternative approach monitoring the effects of individual EBV genes when expressed in an epithelial environment. The target cells of choice were SCC12F, an immortalized non-tumorigenic sub-clone of the human squamous cell carcinoma line SCC12, which retain several characteristics of cultured normal keratinocytes, including responsiveness to terminal differentiation signals^{6,7}. Our interest in the effects of LMP in epithelium stemmed from the findings that (1) the protein is capable of inducing malignant transformation in established rodent fibroblast lines^{8,9}, and (2) it is one of only two EBV proteins detectable in NPC biopsy cells^{2,3}, the other being the nuclear antigen EBNA 1, which is essential for virus genome maintenance¹⁰. SCC12F cells were transfected with the LMP expression plasmid pSVgptMTLM, which allows selection in mycophenolic acid⁸. All 10 drug-resistant clones obtained were LMP-positive by both immunoblotting and immunofluorescence using the specific monoclonal antibodies (mAbs) CS1-4 (ref. 11); SCC12F cells transfected with the control plasmid pSV2gpt served as negative controls in these assays. Figure 1a presents an immunoblot of protein extracts from three representative LMP-positive clones (LMP 4, 7 and 19) illustrating expression of the 66K viral protein (relative molecular mass, 66,000) at levels equal to or greater than those seen in an EBV-transformed B-cell line. By immunofluorescence, the LMP transfectants showed both membrane-associated and cytoplasmic staining with mAbs CS1-4 (Fig. 1b) exactly as do B-cell lines¹¹. Because the effects of LMP in rodent cell transformation assays resemble those produced by activated *ras* proteins,

Received 17 August 1989; accepted 26 February 1990.

1. Chee, M. S. *et al.* *Curr. Topics Microbiol. Immun.* (in the press).
2. Alford, C. A. & Britt, W. J. in *Virology* (ed. Fields, B. N. *et al.*) 629-660 (Raven, New York, 1985).
3. Lefkowitz, R. J. & Caron, M. G. *J. biol. Chem.* **263**, 4993-4996 (1988).
4. Jackson, T. R. *Curr. Opin. Cell Biol.* (in the press).
5. Firtel, R. A., van Haastert, P. J. M., Kimmel, A. R. & Devreotes, P. N. *Cell* **58**, 235-239 (1989).
6. McFarland, K. C. *et al.* *Science* **245**, 494-499 (1989).
7. Lipman, D. J. & Pearson, W. R. *Science* **227**, 1435-1441 (1985).
8. Pearson, W. R. & Lipman, D. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).
9. Staden, R. *CABIOS* **4**, 53-60 (1988).
10. Staden, R. *CABIOS* **5**, 89-96 (1989).
11. McGeoch, D. J. *et al.* *J. gen. Virol.* **69**, 1531-1574 (1988).
12. Davison, A. J. & Scott, J. E. *J. gen. Virol.* **67**, 1759-1816 (1986).
13. Baer, R. *et al.* *Nature* **310**, 207-211 (1984).
14. Ruger, B. *et al.* *J. Virol.* **61**, 446-453 (1987).
15. Mocarski, E. S., Pereira, L. & McCormick, A. L. *J. gen. Virol.* **69**, 2613-2621 (1988).
16. Chang, C.-P., Vesole, D. H., Nelson, J., Oldstone, M. B. A. & Stinski, M. F. *J. Virol.* **63**, 3330-3337 (1989).
17. Chang, C. P., Malone, C. L. & Stinski, M. F. *J. Virol.* **63**, 281-290 (1989).
18. Gretch, D. R., Kari, B., Gehrz, R. C. & Stinski, M. F. *J. Virol.* **62**, 1956-1962 (1988).
19. Gretch, D. R. & Stinski, M. F. *Virology* **174**, 522-532 (1990).
20. Kozak, M. *Nucleic Acids Res.* **9**, 5233-5252 (1981).
21. Weston, K. & Barrell, B. G. *J. molec. Biol.* **192**, 177-208 (1986).
22. Libert, F. *et al.* *Science* **244**, 569-572 (1989).
23. Flesler, S. J. & Basiner, S. F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1116-1120 (1985).
24. Karnik, S. S., Sakmar, T. P., Chen, H.-B. & Khorana, H. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8459-8463 (1988).
25. Dixon, R. A. F. *et al.* *EMBO J.* **6**, 3269-3275 (1987).
26. Brandt, C. J. & Deber, C. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 917-921 (1986).
27. Strader, C. D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 4384-4388 (1987).
28. Dohlman, H. G., Caron, M. G., Strader, C. D., Amlaiki, N. & Lefkowitz, R. J. *Biochemistry* **27**, 1813-1817 (1988).
29. Dixon, R. A. F., Sigal, I. S. & Strader, C. D. *Cold Spring Harb. Symp. quant. Biol.* **53**, 487-497 (1988).
30. Strader, C. D. *et al.* *J. biol. Chem.* **263**, 10267-10271 (1988).
31. Franke, R. R., Sakmar, T. P., Oprian, D. D. & Khorana, H. G. *J. biol. Chem.* **263**, 2119-2122 (1988).
32. Sibley, D. R., Benovic, J. L., Caron, M. G. & Lefkowitz, R. J. *Cell* **48**, 913-922 (1987).
33. Fickenscher, H., Stamminger, T., Ruger, B. & Fleckenstein, B. *J. gen. Virol.* **70**, 107-123 (1989).
34. Hunninghake, G. W., Monick, M. M., Liu, B. & Stinski, M. F. *J. Virol.* **63**, 3026-3033 (1989).
35. Stinski, M. F. *J. Virol.* **23**, 751-767 (1977).
36. Julius, D., Livelli, T. J., Jessell, T. M. & Axel, R. *Science* **244**, 1057-1062 (1989).
37. Barbacid, M. A. *Rev. Biochem.* **56**, 779-827 (1987).
38. Geder, L. in *Oncogenic Herpesviruses*, Vol. 2, 47-60 (ed. Rapp, F.) (CRC, Florida, 1980).
39. Black, J. *Science* **245**, 486-493 (1989).
40. McKinlay, M. A. & Rossmann, M. G. A. *Rev. Pharmac. Toxicol.* **29**, 111-122 (1989).
41. Barton, G. J. & Sternberg, M. J. E. *J. molec. Biol.* **198**, 327-337 (1987).
42. McCaldon, P. & Argos, P. *Proteins* **4**, 99-122 (1988).
43. Oram, J. D. *et al.* *J. gen. Virol.* **59**, 111-129 (1982).
44. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
45. Bankier, A. T., Weston, K. M. & Barrell, B. G. *Meth. Enzym.* **155**, 51-93 (1987).
46. Jahn, G. *et al.* *J. Virol.* **61**, 1358-1367 (1987).
47. Nathans, J. A. *Rev. Neurosci.* **10**, 163-194 (1987).
48. O'Dowd, B. F., Lefkowitz, R. J. & Caron, M. G. A. *Rev. Neurosci.* **12**, 67-83 (1989).
49. Fargn, A. *et al.* *Nature* **335**, 358-360 (1988).
50. Koblik, B. K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 46-50 (1987).
51. Masu, Y. *et al.* *Nature* **329**, 836-838 (1987).
52. Kubo, T. *et al.* *Nature* **323**, 411-416 (1986).
53. Nathans, J. & Hogness, D. S. *Cell* **34**, 807-814 (1983).
54. Young, D., Waitches, G., Birchmeier, C., Fasano, O. & Wigler, M. *Cell* **45**, 711-719 (1986).
55. Jackson, T. R., Blair, L. A. C., Marshall, J., Goedert, M. & Hanley, M. R. *Nature* **335**, 437-440 (1988).
56. Klein, P. S. *et al.* *Science* **241**, 1467-1472 (1988).
57. Hagen, D. C., McCaffrey, G. & Sprague Jr, G. F. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1418-1422 (1986).
58. Stockwell, P. A. *Trends biochem. Sci.* **13**, 322-324 (1988).

ACKNOWLEDGEMENTS. We thank Jon Oram and Peter Greenaway for the AD169 HindIII fragment clones.

* To whom correspondence should be addressed.

SCC12F cells were transfected in parallel experiments with the activated c-Ha-ras expression plasmid pH06T1, yielding stable, high-expressing clones, and with the appropriate control plasmid Homer 6 (ref. 12).

Expression of LMP in the human B-cell environment is associated with the up-regulation of particular cell-surface proteins, notably 'activation antigens' such as CD23 and CD40, and the

cellular adhesion molecules LFA-1, ICAM-1 and LFA-3 (refs 13 and 14). We analysed the LMP-transfected epithelial cells for such surface markers, because NPC cells express high levels of CD40 (ref. 15) and ICAM-1 (P. Busson and L.S.Y., unpublished results). Table 1 presents flow-activated cell sorting (FACS) analysis of the surface phenotype of SCC12F parental cells and of representative transfectants. LMP-expressing clones

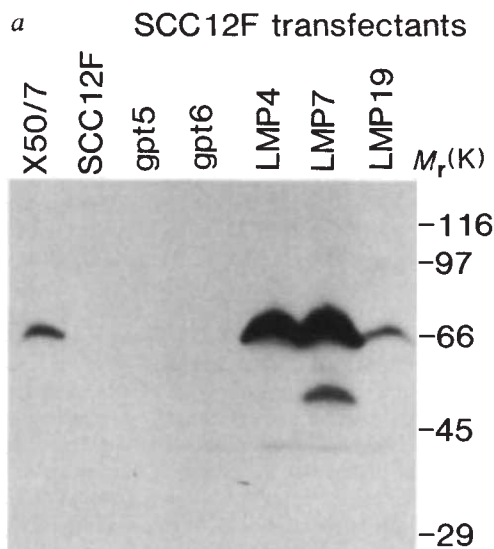
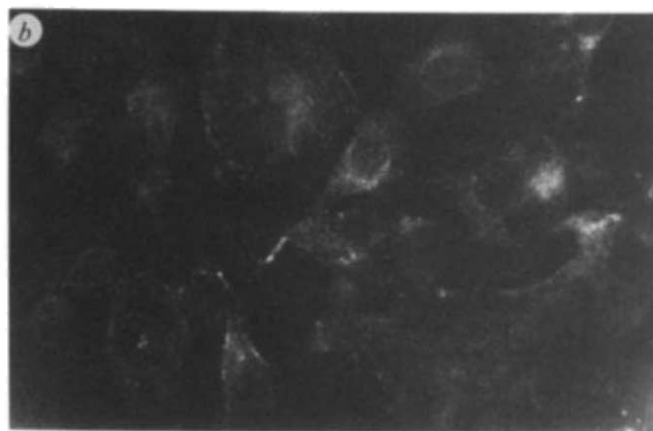


FIG. 1 Expression of LMP in SCC12F transfectants. *a*, Immunoblot showing LMP expression in pSV2gptMTLM transfectants of SCC12F (LMP 4, LMP 7, LMP 19). LMP expression was detected with the anti-LMP mAbs CS1-4, which also cross-react in immunoblots with cellular protein (M_r 41,000) as previously described¹¹. X50/7, an EBV-transformed B-cell line, is a LMP-positive control; SCC12F parent and pSV2gpt transfectants (gpt 5, gpt 6) are negative controls. *b*, Immunofluorescent staining of a pSV2gptMTLM transfectant of SCC12F (LMP 7) with CS1-4 mAbs; no staining was obtained with SCC12F parent and pSV2gpt control transfectants (data not shown). Magnification, $\times 200$.

METHODS. Exponentially growing cultures of SCC12F were rinsed free of the X-irradiated fibroblast feeders with 0.02% EDTA, washed twice in PBS medium, and removed with trypsin: EDTA (0.025: 0.01%). The resulting single cell suspension of SCC12F was recovered in normal growth medium (NGM: DME/F12 3:1, 10% FCS, 2 mM glutamine, $0.4 \mu\text{g ml}^{-1}$ hydrocortisone) and counted. Cells (2×10^6) were resuspended in 1 ml PBS, transferred into an electroporation chamber (Bio-Rad) and $10 \mu\text{g}$ of appropriate plasmid DNA



added. The suspension was left on ice for 10 min and electroporated at 125 μF and 0.45 kV. After a 10 min recovery the cells were washed and counted, and viable cells plated at 5×10^4 cells per 5 cm dish containing 5×10^5 drug-resistant fibroblast feeders. The appropriate drug was added to the medium after 2-3 days (for pSV2gpt and pSV2gptMTLM, $25 \mu\text{g ml}^{-1}$ mycophenolic acid, $160 \mu\text{g ml}^{-1}$ xanthine, $10 \mu\text{g ml}^{-1}$ hypoxanthine; for Homer 6 and pH06T1, $400 \mu\text{g ml}^{-1}$ G418). After ~ 4 weeks of culture, drug-resistant colonies were picked and expanded in drug-containing NGM. Drug-resistant SCC12F clones transfected with pSV2gptMTLM were examined for LMP expression. Protein extracts of the cells were run on 7.5% SDS-PAGE gels, immunoblotted and probed with CS1-4 mAbs as previously described¹¹. For immunofluorescence, cells were plated on teflon-coated microscope slides and grown for 3 days before fixation in methanol at -20°C for 10 min. The slides were stained with CS1-4 mAbs followed by FITC-conjugated anti-mouse immunoglobulin (Sigma) as previously described¹¹; no staining was obtained when pSV2gptMTLM cells were exposed to the FITC-conjugated second step reagent alone.

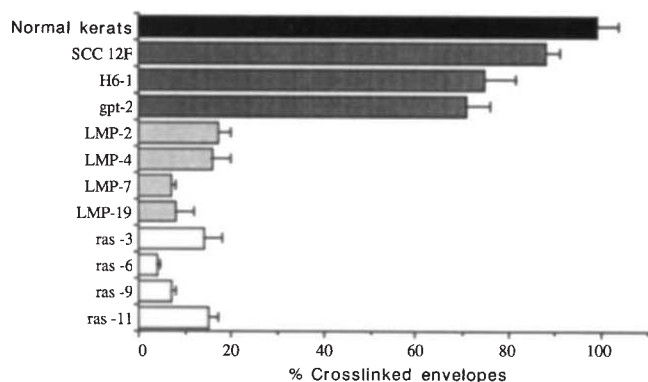


FIG. 2 Ability of SCC12F transfectants to form cross-linked envelopes. Cultures of normal human foreskin keratinocytes, parental SCC12F cells and control transfectants produced using either the Homer 6 (H6-1) or pSV2gpt (gpt-2) plasmid were compared with LMP transfectants (LMP 2, 4, 7, 19) and activated Ha-ras transfectants (ras 3, 6, 9, 11). Data are expressed as the mean ($\pm$ s.e.m.) percentages of ionophore-treated cells that formed cornified envelope structures from assays on triplicate cultures of each of the lines examined. The results represent three separate tests on these particular clones; analysis of a further five control transfectants, three LMP transfectants and three ras transfectants showed the same pattern of results. **METHODS.** All measurements were made on cell cultures grown on 60-mm culture dishes held at confluence for 1 week before analysis. Disaggregated cell suspensions were resuspended at 10^6 cells in 0.5 ml serum-free DME containing $50 \mu\text{g ml}^{-1}$ of A23187 ionophore. After 5 h at 37°C , 0.5 ml of 2% SDS-40 mM DTT was added and the cell suspensions was then removed and scored for numbers of crosslinked envelopes by phase-contrast microscopy as previously described¹⁶.

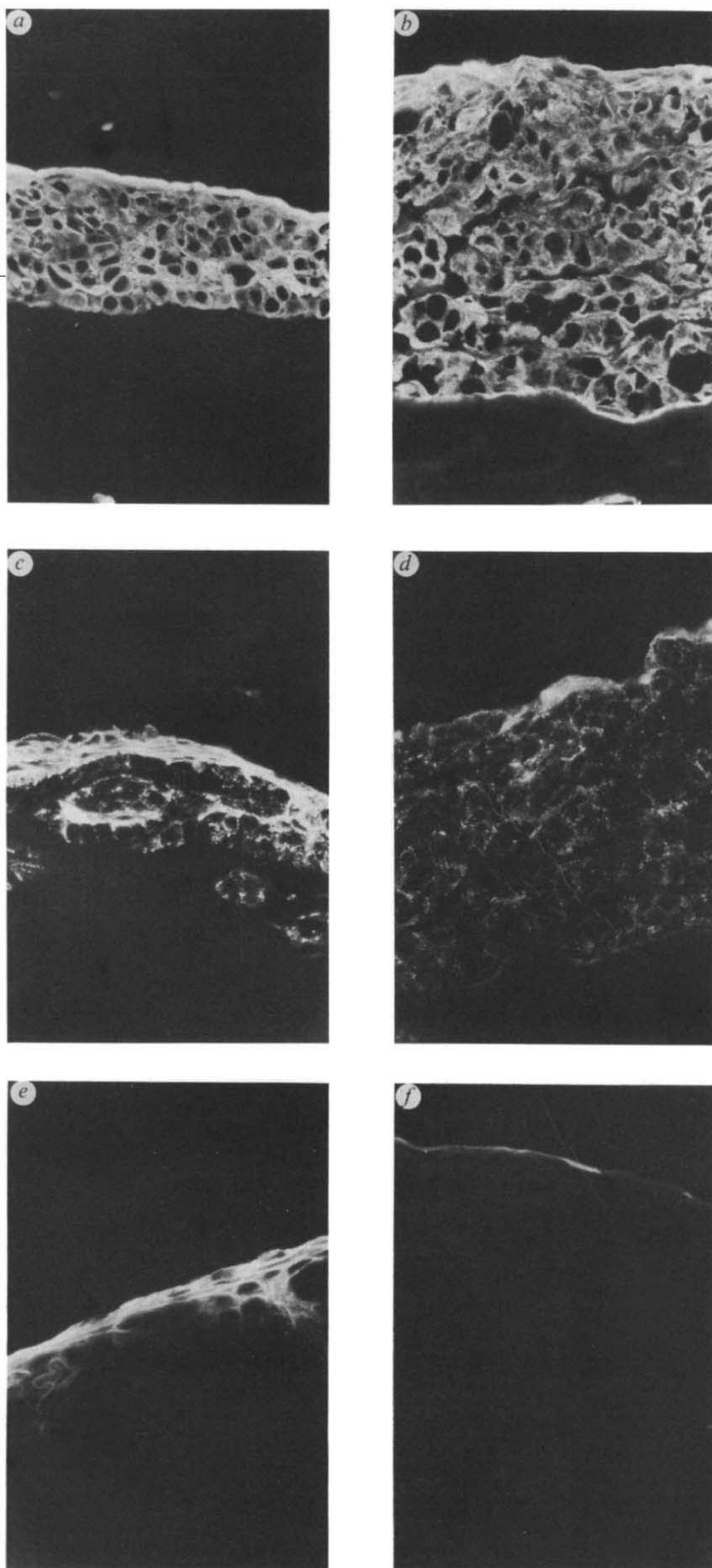


FIG. 3 Immunofluorescent staining of cross-sections of SCC12F transfectants grown on collagen rafts. Staining of sections of control transfectant gpt 2 (a, c, e) and LMP transfectant LMP 19 (b, d, f) using mAb AE1 (ref. 18) against type I keratins (a, b), mAb ZK31 (ref. 28) against desmosomes (c, d) and a specific polyclonal rabbit antiserum²⁹ against involucrin (e, f). The differences observed are representative of those seen with 6 independent transfectants of each type. Magnification, $\times 200$.

METHODS. A collagen solution was prepared using an 8:1:1 mixture of type 1 collagen/10 $\times$ E medium/buffer at 4 $^{\circ}$ C as described by the manufacturers (Seikagaku, Japan). After addition of 1.5×10^5 fibroblast feeders per ml to the collagen solution, 2 ml was added to a 35-mm culture dish, and the collagen allowed to gel at 37 $^{\circ}$ C forming a fibroblast-impregnated collagen raft. SCC12F parent cells or transfectants were then seeded on to the surface of the collagen rafts at 3×10^5 cells per raft, medium was added to submerge the rafts, and the cells were grown to confluence. At confluence, the collagen rafts were raised on to stainless steel grids at the air-medium interface such that subsequent feeding occurred only from below¹⁷. After 3 weeks, the rafts were snap-frozen in liquid nitrogen, cryo-sectioned and the multi-layered epithelium that had formed at the air-medium interface was stained by indirect immunofluorescence, using the above antibodies as a first step and appropriate FITC-labelled anti-mouse immunoglobulin or anti-rabbit immunoglobulin reagents (Sigma) as a second step. Use of the second step reagents alone never gave significant staining.

consistently displayed increased levels of both CD40 and ICAM-1 when compared with parental cells or control transfectants. The CD23 and LFA-1 markers are not naturally expressed on epithelium and were not up-regulated by LMP, whereas LFA-3 expression was not increased above the high level already found on SCC12F parental cells. The *ras* transfectants showed a similar cell surface profile to the LMP transfectants, with the same degree of ICAM-1 up-regulation but with intermediate elevation of CD40 levels (data not shown).

Stably established LMP-expressing clones showed an exponential growth rate in subconfluent culture similar to that of control transfectants and of the parental SCC12F cell line, but significantly lower than that of the *ras* transfectants. Under confluent conditions, however, the LMP-expressing clones resembled *ras*-transfectants in showing little evidence of the cell stratification reproducibly observed in dense cultures of SCC12F and control transfectants. To examine these differences more closely, we first monitored the ability of the various cell clones to form cross-linked envelopes in response to treatment with the calcium ionophore A23187, an established *in vitro* assay reflecting epithelial cell capacity for terminal differentiation¹⁶. Both the LMP and the *ras* transfectants were severely impaired in their ability to form cross-linked envelopes compared with control cells (Fig. 2).

In subsequent experiments, the various transfectants were analysed for stratification on collagen rafts, a culture system that most closely resembles human epithelial cell differentiation *in vivo*¹⁷. In this situation, parental SCC12F cells and six independent control transfectants produced a relatively well organized epithelial structure with a distinct basal cell layer and some degree of stratification in outer layers; in contrast, all of the six LMP-expressing clones analysed produced a much thicker, but less organized epithelium with poorer intercellular contacts. These general features are illustrated in Fig. 3a and b,

where cross-sections of the two types of epithelium have been stained for type I keratins using the mAb AE1 (ref. 18). Specific differences could be observed following staining for desmosomes, a marker of interepithelial cell contact¹⁹. Control transfectants showed patchy staining in the lower, less-differentiated layers, but staining intensified as tight desmosomal junctions were formed during stratification (Fig. 3c). By contrast, the LMP transfectants showed patchy staining throughout the epithelium, with no evidence of tight desmosomal junctions (Fig. 3d). Stratification could also be monitored by staining for involucrin, one of the precursor proteins of the crosslinked envelope²⁰. Control transfectants showed clear evidence of involucrin synthesis in the upper, more differentiated half of the epithelium (Fig. 3e), whereas very few, if any, involucrin-positive cells formed in the LMP transfectant cultures (Fig. 3f). In the same assays, all four *ras* transfectants analysed formed even thicker multi-layer structures than the LMP transfectants, with a similar lack of terminal differentiation markers (data not shown).

Our experimental system has the potential to identify genes whose expression can contribute to the neoplastic transformation of human epithelium. We have shown that LMP can induce changes in the surface phenotype and differentiation responsiveness of human epithelial cells; in certain respects, therefore, LMP appears to be functionally similar to activated *ras* proteins²¹. The significance of the observed surface changes in relation to growth regulation is not known, but it is interesting that CD40, a molecule with homology to nerve growth factor receptor²², is selectively expressed on the proliferating basal layer of normal epithelium and is retained on a variety of carcinomas^{23,24}. More importantly, the ability of LMP to inhibit terminal differentiation in epithelial cells indicates a mechanism whereby EBV infection of stratified nasopharyngeal epithelium could provide an expanded pool of target cells susceptible to the further changes that are necessary for fully malignant transformation to undifferentiated NPC. Such changes might be influenced by other, perhaps novel, EBV gene products^{25,26} and/or by chemical carcinogens²⁷. Our preliminary observation that LMP-transfected SCC12F cells, unlike the highly malignant *ras* transfectants, produce tumours in nude mice only rarely and after a long latency period, suggests that this system could be used to identify such complementary changes. □

TABLE 1 LMP expression and the SCC12F epithelial cell phenotype

Cells	LMP	'B cell associated'		Adhesion molecules		
		CD23	CD40	LFA 1	ICAM1	LFA 3
SCC12F	—	0	45	2	48	102
gpt1	—	1	45	2	47	152
gpt5	—	0	46	0	52	133
LMP19	++	0	89	4	153	103
LMP4	+++	4	114	2	147	137
LMP7	+++	1	134	2	142	122

The results, from one representative experiment, are expressed as mean fluorescence intensity values obtained from cytofluorometric analysis after immunofluorescence staining with mAbs specific for the relevant cell surface markers. Mean fluorescence intensity values are in each case calculated for that percentage of the cell population showing positive (above background) staining and are expressed on an arbitrary linear scale; the values obtained in eight separate analyses of these same cells lay within $\pm 15\%$ of those values presented. LMP expression not only increased the mean fluorescence intensity of CD40-positive and ICAM-1-positive cells in transfected clones (see Table) but also increased the percentage of cells showing positive staining for either marker from 50–75% in SCC12F and control transfectants to >90% in LMP transfectants. These results are illustrative of data obtained from five control transfectants and six LMP transfectants in all. Cultures of drug-resistant SCC12F transfectant clones in exponential growth were washed free of X-irradiated fibroblast feeders and a single cell suspension obtained by gentle trypsinization (0.0125% trypsin, 0.02% EDTA). The cells were allowed to recover in NGM for 1 h and aliquoted into $3-4 \times 10^5$ cells per tube for subsequent staining with appropriate dilution of mAbs to CD23 (mAb MHM6), CD40 (mAb G28.5), LFA 1 (mAb MHM23), ICAM1 (mAb RR1/1) and LFA 3 (mAb TS2/9), as previously described^{13,14}. After 1 h at 4 °C the cells were washed twice and then incubated with FITC-conjugated anti-mouse Ig (Sigma) for an additional 1 h at 4 °C. Finally cells were washed and fixed in 1% formal saline before flow cytofluorometric analysis using a FACS 440 (Becton Dickinson).

Received 7 December, 1989; accepted 6 March 1990.

- Epstein, M. A. & Achong, B. G. (eds) *The Epstein-Barr Virus* (Heinemann, London, 1986).
- Young, L. S. *et al.* *J. gen. Virol.* **69**, 1051–1065 (1988).
- Fahraeus, R. *et al.* *Int. J. Cancer* **42**, 329–338 (1988).
- Sixbey, J. W. *et al.* *Nature* **306**, 480–483 (1983).
- Tomei, L. D., Noyes, I., Blocker, D., Holliday, J. & Glaser, R. *Nature* **329**, 73–75 (1987).
- Rheinwald, J. G. & Beckett, M. A. *Cancer Res.* **41**, 1657–1663 (1981).
- Parkinson, E. K., Pera, M. F., Emmerson, A. & Gorman, P. A. *Carcinogenesis* **5**, 1071–1077 (1984).
- Wang, D., Leibowitz, D. & Kieff, E. *Cell* **43**, 831–840 (1985).
- Baichwal, V. J. & Sugden, B. *Oncogene* **2**, 461–467 (1988).
- Yates, J. L., Warren, N. & Sugden, B. *Nature* **313**, 812–815 (1985).
- Rowe, M. *et al.* *J. gen. Virol.* **68**, 1575–1586 (1987).
- Spandidos, D. A. & Wilkie, N. M. *Nature* **310**, 469–475 (1984).
- Wang, D. *et al.* *J. Virol.* **62**, 4173–4184 (1988).
- Wang, F. *et al.* *J. Virol.* (in the press).
- Bussan, P. *et al.* *Int. J. Cancer* **42**, 599–606 (1988).
- Rubin, A. L. & Rice, R. H. *Cancer Res.* **46**, 2356–2361 (1986).
- Kopan, R., Traska, G. & Fuchs, E. *J. Cell Biol.* **105**, 427–440 (1987).
- Tseng, S. C. G. *et al.* *Cell* **30**, 361–372 (1982).
- Cowin, P. & Garrod, D. R. *Nature* **302**, 148–150 (1983).
- Rice, R. H. & Green, H. *Cell* **18**, 681–694 (1979).
- Parkinson, E. K. *Br. J. Cancer* **52**, 479–493 (1985).
- Stamakovic, I., Clark, E. A. & Seed, B. *EMBO J.* **8**, 1403–1410 (1989).
- Young, L. S., Dawson, C. W., Brown, K. W. & Rickinson, A. B. *Int. J. Cancer* **43**, 786–794 (1989).
- Ledbetter, J. A., Clark, E. A., Norris, N., Shu, G. & Hellstrom, I. in *Leucocyte Typing III: White Cell Differentiation Antigens* (ed. McMichael, A. J.) 432–435 (Oxford University Press, 1987).
- Griffin, B. E. & Karan, L. *Nature* **309**, 78–82 (1984).
- Hitt, M. M. *et al.* *EMBO J.* **8**, 2639–2651 (1989).
- Ito, Y. in *The Epstein-Barr Virus: Recent Advances* (eds Epstein, M. A. & Achong, B. G.) 207–236 (Heinemann, London, 1986).
- Lang, A. B., Odermatt, B. F. & Ruettner, J. R. *Exp. Cell Biol.* **54**, 61–72 (1986).
- Dover, R. & Watt, F. M. *J. invest. Dermat.* **89**, 349–352 (1987).

ACKNOWLEDGEMENTS. We thank Dr E. Kieff for the pSV2gpt and pSV2gpt MTLM plasmids, Dr D. Spandidos for the pH06T1 plasmid and Dr F. Watt for the anti-involucrin antiserum. This work was supported by the Cancer Research Campaign, UK.

Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal

Patrice Lerouge*, Philippe Roche*, Catherine Faucher†, Fabienne Maillet†, Georges Truchet†, Jean Claude Promé* & Jean Dénarié†

* Centre de Recherches de Biochimie et de Génétique Cellulaire, CNRS LP8201, 118 Route de Narbonne, 31062 Toulouse Cedex, France
† Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP27, 31326 Castanet-Tolosan Cedex, France

RHIZOBIA are symbiotic bacteria that elicit the formation on leguminous plants of specialized organs, root nodules, in which they fix nitrogen¹. In various *Rhizobium* species, such as *R. leguminosarum* and *R. meliloti*, common and host-specific nodulation (*nod*) genes have been identified which determine infection and nodulation of specific hosts¹. Common *nodABC* genes^{2–5} as well as host-specific *nodH* and *nodQ* genes^{4,6–8} were shown recently, using bioassays, to be involved in the production of extracellular Nod signals. Using *R. meliloti* strains overproducing symbiotic Nod factors, we have purified the major alfalfa-specific signal, NodRm-1, by gel permeation, ion exchange and C₁₈ reverse-phase high performance liquid chromatography. From mass spectrometry, nuclear magnetic resonance, ³⁵S-labelling and chemical modification studies, NodRm-1 was shown to be a sulphated β-1,4-tetrasaccharide of D-glucosamine (M_r 1,102) in which three amino groups were acetylated and one was acylated with a C₁₆ bis-unsaturated fatty acid. This purified Nod signal specifically elicited root hair deformation on the homologous host when added in nanomolar concentration.

Cytological studies of nodules induced by *R. meliloti* mutants unable to form infection threads or to be released into host cells suggested that the bacteria are able to elicit, at a distance, cortical cell dedifferentiation and initiation of nodule organogenesis by means of a 'nodule-inducing principle' (refs 9, 10). Recent studies showing that alfalfa clones may develop nodules of normal anatomy in the absence of *Rhizobium* indicated that the host plant possesses the entire genetic programme for nodule organogenesis and that the role of the microsymbiont *R. meliloti* is to switch on this programme¹¹. The common *nodABC* and host-specific *nodH* and *nodQ* genes of *R. meliloti*, which are required for specific infection and for triggering nodule formation, determine the production of extracellular symbiotic signals^{1,4,6–8,12,13}.

The transcription of *R. meliloti nod* genes is under the control of *syrM* and three *nodD* regulatory genes, and requires the presence of flavonoids from plant exudates^{1,14,15}. After induction of *nod* gene transcription by the flavonoid luteolin, the sterile filtrate of a *R. meliloti* wild-type strain culture exhibits specific root hair deformation (Had) activity on alfalfa, the homologous host, but not on vetch, a heterologous host⁴. Both the common *nodABC* genes and host-range genes *nodH* and *nodQ* are required for the production of the alfalfa-specific extracellular Nod signals^{4,6–8}. In an attempt to purify these active signal molecules, the filtrate of a luteolin-induced culture of *Rhizobium meliloti* 2011 was extracted by organic solvents (see legend to Fig. 1). Most of the Had activity assayed on alfalfa seedlings was detected in the butanol-soluble and ethyl acetate-insoluble fraction. This extract was fractionated by high performance liquid chromatography (HPLC) on a C₁₈ reverse-phase column.

However, the amount of compounds present in the fraction showing the highest Had activity was too small for structural analysis. To improve the bacterial production of these compounds we introduced the plasmid pGMI149 into *R. meliloti* to provide extra copies of the nodulation region. The plasmid pGMI149 carries the whole set of common and specific *nod* genes as well as three regulatory genes, *nodD1*, *nodD3* and *syrM*, cloned in an Inc-P1 vector present at 5–10 copies per cell^{11,12,15} (Fig. 1a). The introduction of pGMI149 into *R. meliloti* 2011 resulted in both an increase of more than a 100-fold of the Had activity on alfalfa, as estimated by assaying serial dilutions of extract, and the appearance of two far-UV (220 nm) absorbing peaks on HPLC chromatograms (Fig. 1b).

To ascertain the role of the *nodABC* genes in the production of these compounds, *nodA::Tn5* or *nodC::Tn5* insertions were introduced into pGMI149. To avoid interference between the *nod* genes present on the pSym megaplasmid and those present on pGMI149, we introduced pGMI149 and its mutant derivatives either into *R. meliloti* strain GMI766, which carries a deletion of the whole pSym *nod* region, or into GMI6059, a strain carrying a deletion of the *nodD1ABC* genes. The inactivation of the *nodABC* operon by a *nodA::Tn5* insertion resulted simultaneously in the disappearance of the two major UV-absorbing peaks in the HPLC profiles and in the loss of Had activity on alfalfa (Fig. 1b). Inactivation of *nodC* also resulted in no accumulation of these factors. Insertion of Tn5 downstream of *nodC* in the putative *nodIJ* region¹ did not suppress NodRm production, showing that *nodC::Tn5* insertion was not acting by a polar effect on downstream genes (data not shown). Thus the ultraviolet-detected compounds are symbiotically active molecules and the *nodC* gene is required for their production.

To facilitate large-scale culture filtration, the pGMI149 plasmid was introduced into the *R. meliloti exoB* mutant EJ355 which is deficient for exopolysaccharide production¹⁰. The Exo[−] EJ355 (pGMI149) strain exhibited the same Nod factor peaks in HPLC profile as the Exo⁺ strains 2011 (pGMI149) and GMI766 (pGMI149). The filtrates of fermenter cultures of the non-mucoid EJ355 (pGMI149) strain were extracted and successively purified by preparative reverse-phase C₁₈ HPLC, by gel permeation on a Sephadex LH20 column and by ion exchange chromatography on a DEAE column as described in Fig. 1. Only one peak absorbing at 220 nm could be observed which showed Had activity on alfalfa. Analytical HPLC on a C₁₈ reverse-phase column resolved the active fraction into two close UV-absorbing peaks which showed Had activities on alfalfa but not on vetch (Fig. 1b). Ten litres of induced rhizobial culture yielded around 4 mg of purified Nod factors.

Chemical analysis of the two Nod factors was carried out using either each peak or a mixture of both, and this revealed that the two compounds correspond to the α and β anomers of the same molecule, at the C-1 position of the reducing end sugar. We propose calling this molecule NodRm-1. A clear separation of α and β anomers of N-acetyl-D-glucosamine oligomers by reverse-phase HPLC has already been reported¹⁶. The structure of NodRm-1 was determined by mass spectrometry (Fig. 2a, b), chemical modification (Fig. 2) and NMR spectroscopy (Fig. 3). The presence of a sulphate group was confirmed by ³⁵S-labelling (Fig. 1c). These data, together with methylation analysis, led to the structure proposed in Fig. 4. NodRm-1 is an N-acyl-tri-N-acetyl-β-1,4-D-glucosamine tetrasaccharide bearing a sulphate group on carbon 6 of the reducing sugar moiety. The aliphatic chain is carried by the non-reducing terminal sugar and is a 2,9-hexadecadienoic N-acyl group. The rationale for the structural assignments is given in the legends of Figs. 2 and 3. Minor compounds were detected which differed slightly from NodRm-1 in the length of the aliphatic chain (C₁₆, C₁₈) and number and location of double bonds. These molecules were designated NodRm factors since they could not be detected in the extracts of *nodA* or *nodC* mutants. Chemical

and biological studies of the various NodRm factors produced by mutants altered in different specific *nod* genes should indicate whether these minor molecules are intermediates in the biosynthetic pathway of the NodRm-1 symbiotic signal. NodRm-1 was also clearly the major Nod factor in filtrates of Exo⁺ strains 2011 (pGMI149) and 766 (pGMI149), as determined by mass and NMR spectrometry and alfalfa Had activity. Breaking glycosidic bonds by methanolysis resulted in a complete loss of Had activity, as assayed on alfalfa seedlings at concentrations equivalent to 10^{-8} – 10^{-11} M NodRm-1.

The following evidence shows that the NodRm factor described is a plant-specific symbiotic signal: (1) After three types of purification based on different physicochemical properties, ion exchange, gel filtration and partition chromatography, we always observed an absolute correlation between the presence of this molecule (characterized by HPLC profile, mass spectrometry and ¹H-NMR spectroscopy) and specific Had activity on alfalfa. (2) An increase in *nod* gene activity, either by induction of *nod* gene transcription or by gene dosage, was correlated with an increased production of this molecule, whereas a Tn5 mutation in *nodA* or *nodC* genes resulted in no

production (see legend to Fig. 1). (3) The biological activity of NodRm, as detected by the hair-branching bioassay, was very high and specific. A NodRm solution in the 10^{-8} – 10^{-11} M range elicited Had reactions on the homologous host alfalfa but not on vetch, a heterologous host (data not shown). The NodRm factor was still clearly active on alfalfa at a 10^{-11} M concentration. NMR studies indicated that NodRm-1 was at least 95% pure. Molecules having aromatic rings, like the plant hormones auxin and cytokinin, could not be detected. The study of ultraviolet absorbance of NodRm-1 at 260 and 280 nm showed that if such hormone-like compounds were present in the active fraction, they should represent less than 1% of the purified product. Therefore any possible hormone-like contaminants would be below 10^{-13} M in our assays, that is, at a concentration around a million-fold below the level (10^{-7} M) at which these phytohormones act when added exogenously¹⁷. Indole acetic acid, isopentyl adenine and zeatin did not comigrate with NodRm-1 and did not exhibit any Had activity on alfalfa in the range 10^{-6} – 10^{-13} M. We thus conclude that such contaminants are very unlikely to account for the observed effects.

In Gram-negative bacteria *N*-acetyl-D-glucosamine is in-

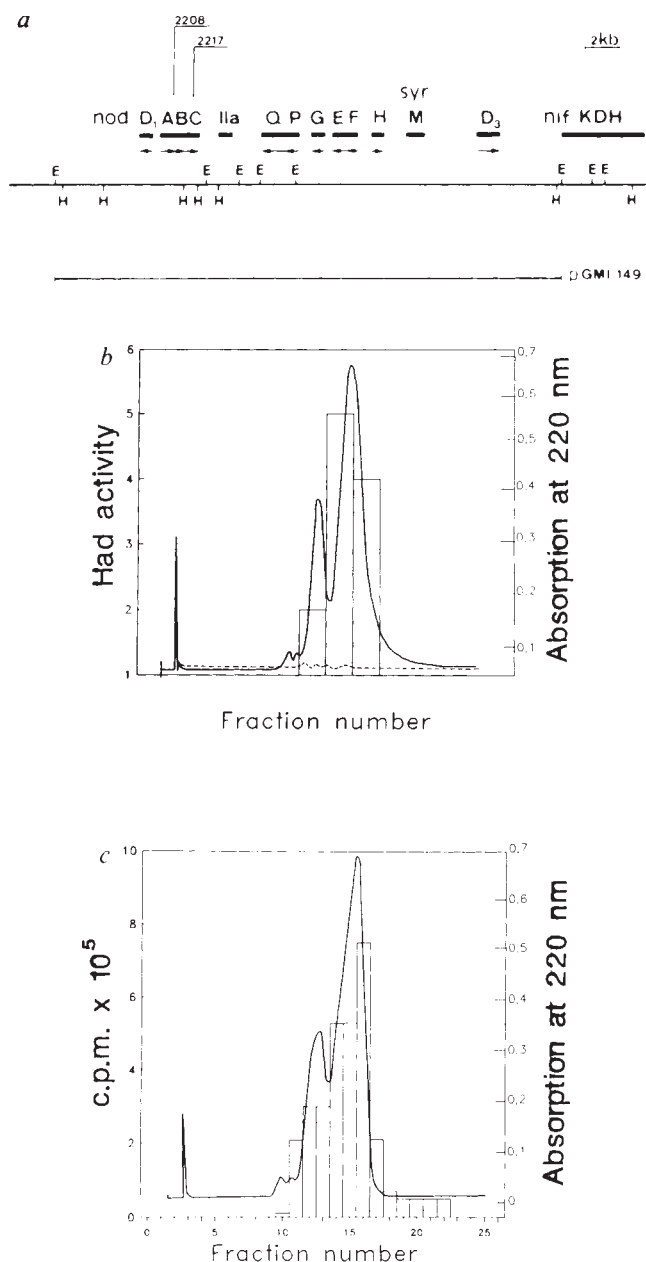


FIG. 1 *a*, Physical and genetic map of the *nod* region of *R. meliloti* 2011 (refs 1, 4, 13, 15). The restriction map (E, EcoRI; H, HindIII) is shown with the pGMI149 plasmid represented below¹². Arrows indicate the direction of transcription of the *nod* genes. The position of *nodA*::Tn5 (2,208) and *nodC*::Tn5 (2,217) mutations is given above the map¹². The putative *nodJ* genes are located downstream of *nodABC* (ref. 1). *b*, Analytical HPLC chromatogram (isocratic conditions) of butanol-extracted compounds from a filtrate of *R. meliloti* EJ355 (pGMI149), a NodRm factor-overproducing strain. The elution profile of a sample of 300 µg of NodRm-1, as monitored by absorption at 220 nm, is represented by a solid line. The fraction size was 3 ml. The hair-branching activity on alfalfa is shown as a bar graph. Similar results were obtained with two other overproducing strains 2011 (pGMI149) and GMI766 (pGMI149). In contrast, for strains having a *nod* deletion in the pSym (GMI766 and GMI6059) and a Tn5 insertion in the *nodA* or *nodC* genes located on pGMI149, no peaks of far ultraviolet-absorption (dotted line) and no Had biological activity could be detected. *c*, Demonstration of the presence of a sulphate group in NodRm factors by labelling with [³⁵S]sodium sulphate. Analytical HPLC chromatogram run in the same conditions as for *b*. The bar graph represents ³⁵S radioactivity. The radioactive peak co-migrated with the UV-absorbing peak. In GMI6059 (pGMI149*nodA*::Tn5), a *NodA*⁻ derivative, no peak of ³⁵S radioactivity could be detected.

METHODS. *R. meliloti* EJ355 is an Exo⁻ derivative of strain 2011 (ref. 10). GMI766 carries a deletion of the whole *nod* region¹². GMI6059 was constructed by introducing a *nodD1ABC* deletion¹² into EJ355 by marker exchange. pGMI149, pGMI149*nodA*::Tn5 or pGMI149*nodC*::Tn5 were introduced into GMI766 and GMI6059 as already described¹². Bacteria were grown in a liquid minimal medium²³ containing Na glutamate (1 g l⁻¹) and Na succinate (2 g l⁻¹) as nitrogen and carbon sources, and luteolin (10 µM) as a *nod* gene inducer¹⁴. Log phase cultures (~5 × 10⁸ colony-forming units per ml) were centrifuged, sterilized through filter membranes (0.45 µm) and extracted by butanol. The residue was dissolved in water and extracted with ethyl acetate. The water fraction was lyophilized and chromatographed successively as follows: (1) on a preparative reverse-phase C₁₈ column (10 µm, 7.5 × 250 mm) in a linear water-acetonitrile gradient (solvent x = 80/20, solvent y = 50/50) at a flow rate of 2 ml min⁻¹; (2) on a Sephadex LH20 column (99% ethanol); (3) on a DEAE trisacryl column equilibrated with Tris-HCl (10⁻³ M, pH8) and eluted with a linear gradient of sodium chloride (up to 0.1 M); (4) the final purification of the active fraction was achieved with an analytical reverse-phase C₁₈ column (5 µm, 4.6 × 250 mm) in isocratic conditions (water-acetonitrile 80/20) at a flow rate of 1 ml min⁻¹. NodRm was labelled by growing bacteria overnight with 80 µCi of ³⁵S-sodium sulphate. Incorporation was assayed by liquid scintillation counting. The biological Had activity was assayed microscopically by observing root hair deformation on alfalfa⁷ (*Medicago sativa*) and on common vetch³ (*Vicia sativa* subsp. *nigra*). Fractions were serially diluted and a minimum of 10 plants were observed for each dilution. Intensity of the Had reaction was scored from 0 to 6 as a function of both the proportion of branched root hairs and the extent of the root region affected. To avoid subjective bias during visual rating, samples were scored without previous knowledge as to the specific treatment used. Had results given in this figure were obtained with dilutions corresponding to around 10⁻¹⁰ M NodRm-1 factor.

volved in the biosynthesis of two major components of the cell wall, peptidoglycan in the periplasmic space and lipidA in the outer membrane¹⁸. Future study of the biosynthesis of Nod signals will address the question of whether the *nod* gene products mediate a new synthetic pathway from *N*-acetyl-D-glucosamine or, rather, if they are involved in the degradation of pre-existing cell wall macromolecules. Some fungal oligosac-

charide elicitors, which trigger host defence mechanisms such as phytoalexin production, have been shown to be derived from fungal cell wall degradation and are active in the nanomolar range¹⁹. It is worth noting that NodRm is a plant-specific signal, in contrast to the purified fungal oligosaccharide elicitors so far described which are not host species-specific¹⁹.

In addition to the common *nodABC* genes, the host-range

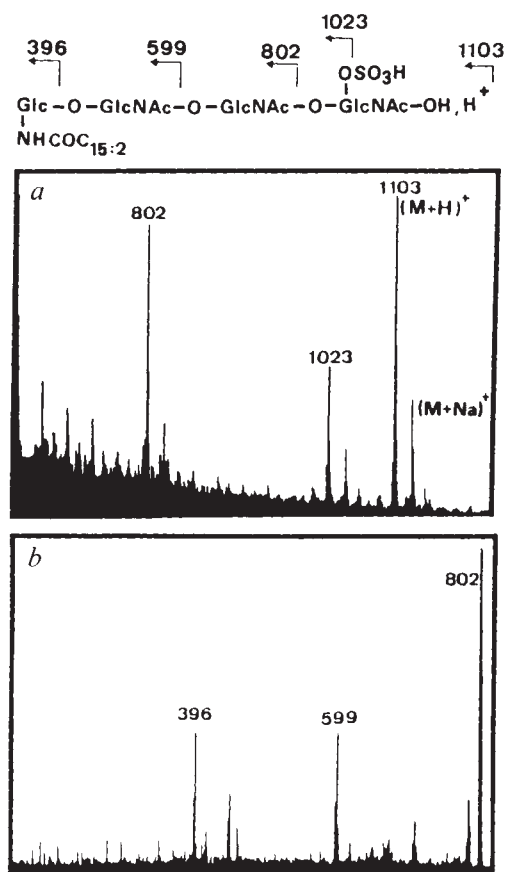
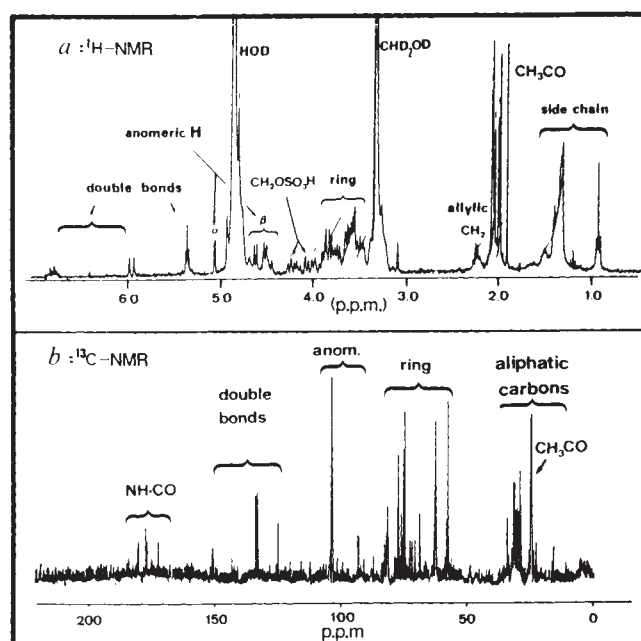


FIG. 3 NMR spectrometry of NodRm-1. *a*, The ^1H NMR spectrum (CD_3OD ; Bruker AM 300 MHz) showed four olefinic protons which were assigned to the internal double bond of the acyl chain ($\delta=5.34$ p.p.m.; 2H) and to the E conjugated double bond ($\delta=5.95$ p.p.m.; d; $J=15$ Hz and $\delta=6.83$ p.p.m.; dt; $J=7$ and 15 Hz). The structure of the acyl chain was confirmed by ^1H NMR-COSY experiments (data not shown). The proton anomeric region revealed three β -interglycosidic linkages ($J=8.5$ Hz). The α ($J=3$ Hz) and β ($J=8.5$ Hz) anomeric protons of the terminal reducing *N*-acetylglucosamine were also observed, and they disappeared after NodRm-1 reduction ($\text{NaBD}_4/\text{H}_2\text{O}$; 40 °C; 2 h). The 3.4 to 3.9 p.p.m. region was assigned to the sugar ring protons²⁹. Two downfield signals ($\delta=4.07$ and 4.26 p.p.m.; dd) were attributed to two H-6 protons allowing the location of the sulphate group at C-6 (see ref. 30). Six methyl signals (6×1.5 H) were assigned to the three *N*-acetyl groups of each α and β anomer of native NodRm-1. *b*, All the signals observed in the ^{13}C NMR spectrum (D_2O ; ref: TMS; Bruker AC 200 MHz) have been assigned to the different carbons of NodRm-1, in agreement with previous reports^{29,30}. This spectrum confirmed the $\beta(1 \rightarrow 4)$ *N*-acetylglucosamine tetrasaccharide structure ($\delta \text{ C1}=103.9$ p.p.m. and $\delta \text{ C4}=81.8\text{--}82.0$ p.p.m.), the presence of a bis-unsaturated chain and a sulphate substitution characterized by the downfield shift of one C-6 of a glucosamine residue ($\delta=68.8$ p.p.m.).

FIG. 2 Mass spectrometry (VG Analytical, ZAB 2E) and analysis by chemical degradation of NodRm-1. *a*, Fast atom bombardment (FAB-MS) spectra were obtained through a magnetic scan from 4,000 to 100 mass units. NodRm-1 exhibited a pseudomolecular ion ($\text{M}+\text{H}^+$) at $m/z=1,103$ by positive ion MS and a ($\text{M}-\text{H}^-$) ion at $m/z=1,101$ in the negative mode. A fragment was observed at $m/z=1,023$ in the positive spectrum. It arose from the decomposition of the ($\text{M}+\text{H}^+$) ion as confirmed by the metastable fragmentation (B/E linked scan) of the latter. The 80 mass unit loss suggested either a phosphate or a sulphate group²⁴. The latter was confirmed by ^{35}S -labelling. Furthermore, no signal could be detected in ^{31}P NMR spectrometry. Sodium borodeuteride-reduced NodRm-1 ($\text{NaBD}_4/\text{H}_2\text{O}$; 40 °C; 2 h) showed a shift of three mass units on both ($\text{M}+\text{H}^+$) and ($\text{M}+\text{H}^+$) minus 80. However the fragment ion at $m/z=802$, due to the loss of sulphated *N*-acetylglucosamine, was not shifted. This indicated that the sulphate group was carried by the sugar moiety at the reducing end. *b*, Collision-induced decomposition of $m/z=802$ (B/E linked scan) showing the successive elimination of two *N*-acetyl glucosaminyl units by glycosidic cleavages²⁵ (losses of 203 mass units).

METHODS. Chemical analysis of NodRm-1: Acidic hydrolysis (3N HCl; 100 °C; 3 h) released glucosamine which was of the D series as determined by gas chromatography (GC) analysis of its glycoside with (–)-2-butanol (peracetylated derivative). A milder acidic cleavage using 1 N HCl in methanol (80 °C; 1 h) afforded both *N*-acetyl glucosamine and a long chain *N*-acylated glucosamine, the structure of which was confirmed by electron impact-mass spectrometry (EI-MS) of its trimethylsilyl derivative²⁶. The main characteristic fragments were at $m/z=365$ (ion containing the C2–C3 part of the sugar ring) and at $m/z=452$ (C2–C3–C4 part), indicating a bis-unsaturated C_{16} *N*-linked fatty acid. The *N*-linked fatty acid was liberated from NodRm-1 by saponification and analysed by gas chromatography coupled to tandem mass spectrometry (negative ion GC/MS/MS) of its perfluoro-benzyl-ester derivative, inducing remote charge fragmentation processes²⁷. This characterized the major $\text{C}_{16:2}$ fatty acid as bis-unsaturated on both positions 2 and 9 with no chain branching. Other minor C_{16} and C_{18} fatty acids were also detected (less than 5%). The interglycosidic linkage mode of the different glucosamine moieties was determined by methylation analysis of reduced NodRm-1 (ref. 28). The partially methylated alditol acetates, obtained after subsequent chemical reactions, were characterized by EI/GC/MS and this indicated (1 → 4) linkages between all monosaccharidic moieties.



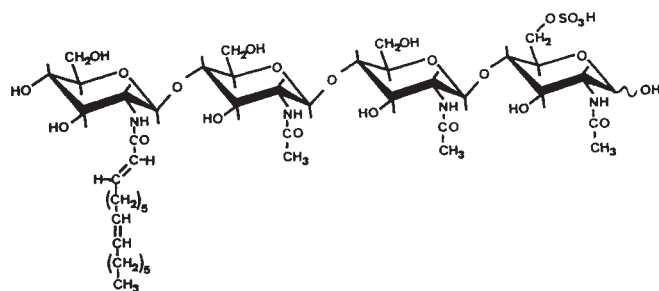


FIG. 4 The proposed NodRm-1 structure is consistent with NMR spectroscopy, mass spectrometry, ^{35}S -labelling and methylation analysis data.

nodH and *nodQ* genes are required for the production of alfalfa-specific signals^{4,6-8}. These *nod* genes are involved in the control of host-specific steps such as recognition, infection and nodulation^{1,12,13}. Studies of host-range *nod* mutants of *R. meliloti* revealed an absolute correlation between the specificity of the symbiotic behaviour of living rhizobial cells and the Had bioassay specificity of their sterile supernatants, which indicates that the Nod signals are involved in determining specific infection and nodulation^{4,7}. We have established here that NodRm is a symbiotic signal which is highly plant-specific. The transfer of a pea lectin gene into white clover seedlings recently achieved resulted in the transfer of the ability to be infected and nodulated by a pea-specific *Rhizobium* (ref. 20). This finding supports the hypothesis that plant lectins are involved in the *Rhizobium*-legume specific recognition. The alfalfa-specific NodRm-1 signal has structural properties which are characteristic of effective lectin ligands such as hexosamine oligosaccharides²¹. A plant root lectin located at the surface of the cytoplasmic membrane of the susceptible root hairs might be part of the Nod signal plant receptor complex²². Purification and characterization of *Rhizobium meliloti* symbiotic signals open up new ways for studying the molecular basis of plant-microbe specific recognition and the mechanisms by which symbiotic microbes signal to plants to elicit various host reactions, including the induction of organogenesis, without triggering plant defence reactions. □

Received 28 December 1989; accepted 2 March 1990.

- Long, S. R. *Cell* **56**, 203-214 (1989).
- van Brussel, A. A. N. et al. *J. Bact.* **165**, 517-522 (1986).
- Zaat, S. A. J., van Brussel, A. A. N., Tak, T., Pees, E. E. & Lugtenberg, B. J. J. *J. Bact.* **169**, 3388-3391 (1987).
- Faucher, C. et al. *J. Bact.* **170**, 5489-5499 (1988).
- Schmidt, J., Wingender, R., John, M., Wieneke, U. & Schell, J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8578-8582 (1988).
- Banfalvi, Z. & Kondoros, A. *Plant molec. Biol.* **13**, 1-12 (1989).
- Faucher, C. et al. *Molec. Plant-Microbe Interact.* **2**, 291-300 (1989).
- Faucher, C. et al. in *Signal Molecules in Plant and Plant-Microbe Interactions* (ed. Lugtenberg, B. J. J.) 379-386 (Springer-Verlag, Berlin/Heidelberg, 1989).
- Truchet, G., Michel, M. & Dénarié, J. *Differentiation* **16**, 163-173 (1980).
- Finan, T. M. et al. *Cell* **40**, 869-877 (1985).
- Truchet, G. et al. *Molec. gen. Genet.* **219**, 65-68 (1989).
- Debellé, F. et al. *J. Bact.* **168**, 1075-1086 (1986).
- Cervantes, E. et al. *Molec. Microb.* **3**, 745-755 (1989).
- Peters, N. K., Frost, J. W. & Long, S. R. *Science* **233**, 977-980 (1986).
- Mulligan, J. T. & Long, S. R. *Genetics* **122**, 7-18 (1989).
- Blumberg, K., Linier, F., Pustilnik, L. & Bush, C. A. *Analyt. Biochem.* **119**, 407-412 (1982).
- Zeroni, M. & Hall, M. A. in *Hormonal Regulation of Development I, Molecular Aspects* (ed. MacMillan, J.) 511-586 (Springer-Verlag, Berlin/Heidelberg, 1980).
- Raetz, C. R. H. in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 498-503 (Am. Soc. Microbiol., Washington, 1987).
- Darvill, A. G. & Albersheim, P. A. *Rev. Pl. Physiol.* **35**, 243-298 (1984).
- Diaz, C. L., Melchers, L. S., Hooykaas, P. J. J., Lugtenberg, B. J. J. & Kijne, J. W. *Nature* **338**, 579-581 (1989).
- Lis, H. & Sharon, N. *Rev. Biochem.* **55**, 35-67 (1986).
- Kijne, J. W., Diaz, C. L. & Lugtenberg, B. J. J. in *Signal Molecules in Plant and Plant-Microbe Interactions* (ed. Lugtenberg, B. J. J.) 351-358 (Springer, Berlin/Heidelberg, 1989).
- Vincent, J. M. *A Manual for the Practical Study of Root-Nodule Bacteria*. IBP Handbook no. 15 (Blackwell Scientific Publications, Oxford, 1970).
- Dell, A. et al. *Carbohydr. Res.* **179**, 7-19 (1988).
- Dell, A. *Adv. Carbohydr. Chem. Biochem.* **45**, 19-72 (1987).

- Demary, M., Puzo, G. & Asselineau, J. *Nouv. J. Chimie* **2**, 373-378 (1977).
- Promé, J. C., Aurelle, H., Couderc, F. & Savagnac, A. *Rapid Commun. Mass Spectrom.* **1**, 80-82 (1987).
- Finne, J., Krusius, T. & Rauvala, H. *Carbohydr. Res.* **80**, 336-339 (1980).
- Boyd, J., Porteous, R. & Soffe, N. *Carbohydr. Res.* **139**, 35-46 (1985).
- Strecker, G., Wieruszski, J. M., Martel, C. & Montreuil, J. *Carbohydr. Res.* **185**, 1-13 (1989).

ACKNOWLEDGEMENTS. We thank C. Pontus (Sanofi Recherche, Toulouse) for mass spectrometry measurements and M. Antoine (Sanofi Elf Biorecherches, Labège) for his help in large-scale culture filtration. We are grateful to A. Stanford and D. Kahn for reviewing the manuscript and to P. Servin (Université Paul Sabatier, Toulouse), F. Pratbernou, F. Couderc and J. Roussel (CRBGC-CNRS, Toulouse) for their help in physicochemical measurements. The two first authors have made equal contributions to this work.

Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins

Bryan F. Anderson, Heather M. Baker, Gillian E. Norris, Sylvia V. Rumball & Edward N. Baker*

Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand

PROTEINS of the transferrin family, which contains serum transferrin and lactoferrin, control iron levels in higher animals through their very tight ($K_{\text{app}} \sim 10^{20}$) but reversible binding of iron^{1,2}. These bilobate molecules^{3,4} have two binding sites, one per lobe, each housing one Fe^{3+} and the synergistic CO_3^{2-} ion⁵. Crystallographic studies of human lactoferrin^{4,6} and rabbit serum transferrin⁷ in their iron-bound forms have characterized their binding sites and protein structure. Physical studies^{8,9} show that a substantial conformational change accompanies iron binding and release. We have addressed this phenomenon through crystal structure analysis of human apolactoferrin at 2.8 Å resolution. In this structure the N-lobe binding cleft is wide open, following a domain rotation of 53°, mediated by the pivoting of two helices and flexing of two interdomain polypeptide strands. Remarkably, the C-lobe cleft is closed, but unliganded. These observations have implications for transferrin function and for binding proteins in general.

High-quality crystals of human apolactoferrin were only obtained after enzymatic removal of the carbohydrate¹⁰. Deglycosylated lactoferrin is, however, indistinguishable from the native protein with respect to iron binding and release (H.M.B., G.E.N. and E.N.B., submitted). The structure was solved by molecular replacement using diffractometer data to 2.8 Å resolution. Search models were based on the iron-lactoferrin structure but as iron loss was expected to result in domain movement, a variety of models were tried: the whole molecule, the individual lobes, and a 'core' comprising the N1 and C1 domains (Fig. 1). All except the N-lobe gave the same correct rotation solution, but the whole molecule was required for a correct translation solution. Initial rigid-body refinement methods accurately positioned three out of the four domains, but the N2 domain could only be placed by rotating it onto weak electron density recognized as three strands of its central β -sheet. The correctness of this model was shown by its subsequent refinement, and the appearance of areas of density representing side chains which had not been included in the search models (being poorly defined in the iron-lactoferrin structure). Further details of the protein conformation and structure solution are given in the legends to Figs 1 and 2, respectively. This structure, after restrained least-squares refinement, has a 0.020 Å r.m.s. deviation from standard bond lengths and the crystallographic R-factor is 0.213 for data between 10-2.8 Å spacing.

* To whom all correspondence should be addressed.

Two remarkable features are apparent in the apolactoferrin structure (Fig. 1). The binding cleft in the N-lobe is wide open, following a rotation of 53° of the N2 domain relative to the N1 domain. In the C-lobe, however, the cleft is closed, although no metal ion is present. This 'one open, one closed' structure was completely unexpected, and poses intriguing questions about the binding process and conformational flexibility.

The open structure of the N-lobe should model the reactive binding form of transferrins, although the extent of conforma-

tional change may vary in different transferrins (see Fig. 3). In lactoferrin, the conformational change has elements of both the simple hinge seen in immunoglobulin structures¹¹ and the helix-mediated domain closures seen in some enzymes¹². First, it involves, as anticipated¹³, the flexing of two anti-parallel extended strands of polypeptide which run behind the iron site connecting the N1 and N2 domains (Fig. 3). The hinge point is around residues His 91 and Pro 251. Second, two helices are in contact in the domain interface; helix 11 (residues 321–332)

FIG. 1 Stereo C_α plot of apo-lactoferrin (yellow) superimposed on iron-lactoferrin (blue). The latter is as described^{4,6} but after further refinement at 2.1 Å resolution (crystallographic R -factor, 0.178). Each molecule is folded into two globular lobes, representing the N-terminal (upper) and C-terminal (lower) halves. The binding cleft in each lobe is between two domains, N1 (upper left) and N2 (upper right) for the N-lobe, and C1 (lower right) and C2 (lower left) for the C-lobe. Here, domains N1, C1 and C2 of apolactoferrin have been superimposed on the corresponding domains of iron-lactoferrin, to demonstrate the close correspondence of the C-lobes and the large movement of the N2 domain. The disulphide bridge, 483–677, which may restrict movement of the C2 domain is shown in orange (centre left).

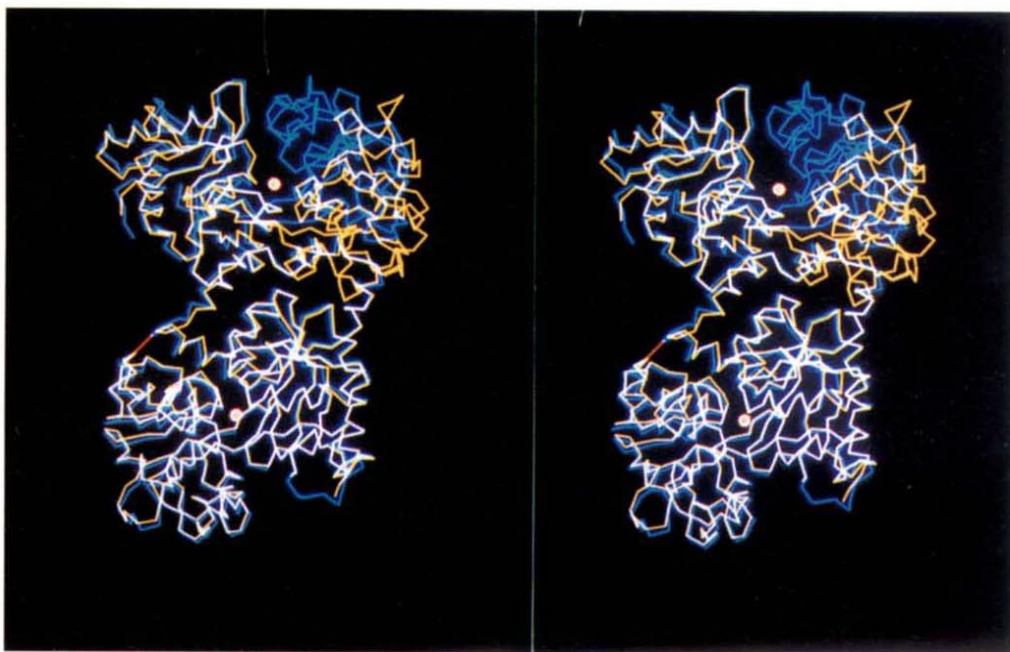
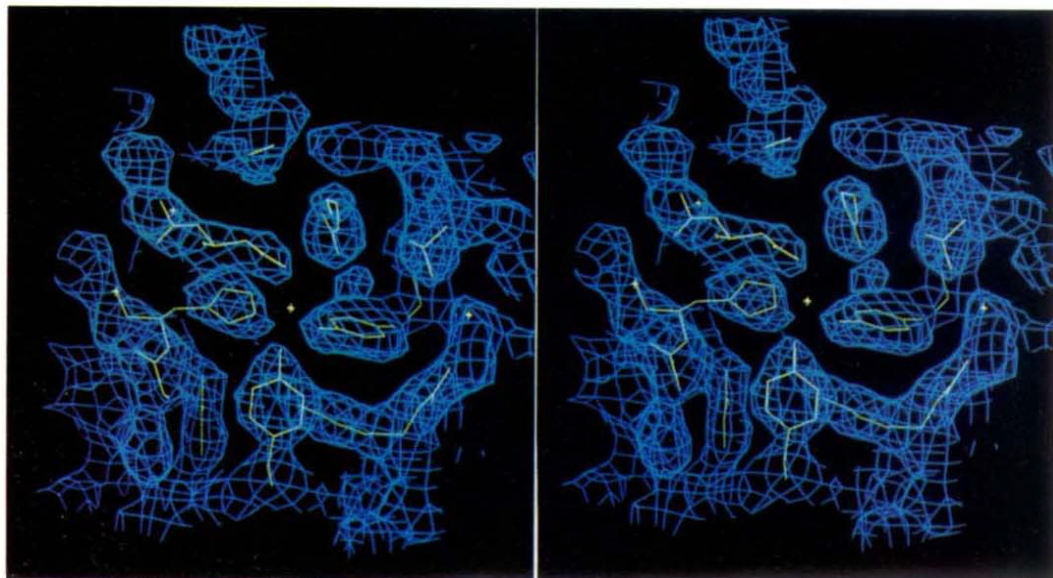


FIG. 2 Electron density in the binding cleft of the C-lobe. The absence of density between the four ligands (clockwise from right, Tyr 528, Tyr 435, His 597, Asp 395) shows no metal is bound. The position occupied by the iron atom in iron-lactoferrin is shown with a cross. A small peak above the ring of Tyr 528 may be a solvent molecule in the anion binding site⁶ (Arg 465 lies above it, and Thr 461 and the helix 5 N-terminus to the right). The map is a $2F_{\text{obs}} - F_{\text{calc}}$ omit map, for which the ligand residues had been omitted from least-squares refinement and phase calculations before calculation of the map.



METHODS. Crystals prepared as described¹⁰. Data to 2.8 Å collected on a CAD4 diffractometer at room temperature. Molecular replacement solution used fast rotation function²² and translational R -factor search. In the initial rigid-body refinement, using the program CORELS²³, the structure was broken into successively smaller blocks (whole molecule; two lobes; four domains; 14 coherent structural elements) as the resolution was increased from 8 Å to 5 Å. After placement of the N2 domain (see text), refinement was by restrained least-squares (program TNT²⁴) with restraints on bond lengths and angles,

planes and non-bonded contacts. After 25 cycles, the model was rebuilt from $2F_{\text{obs}} - F_{\text{calc}}$ omit maps, using the molecular modelling program FRODO²⁵. After 40 more cycles, with individual thermal parameters in the last five, the crystallographic R -factor was 0.213 for 17686 reflections ($I > 0.5\sigma_I$) in the resolution range 10–2.8 Å. The final model has 5,231 protein atoms (residues 3–418, 424–691) with r.m.s. deviations from standard values 0.020 Å (bond lengths) and 3.8° (bond angles).

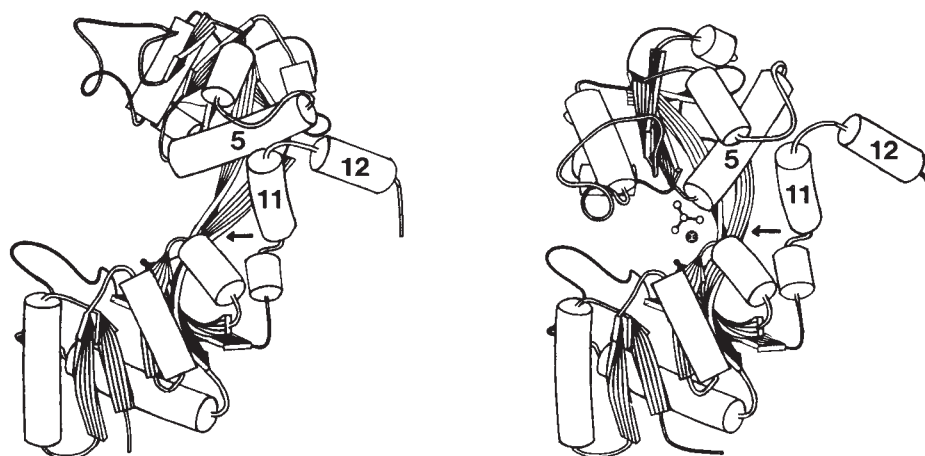


FIG. 3 Schematic representation of the N-lobe polypeptide chain fold, showing the conformational change between open (left) and closed (right) forms. An arrow marks the approximate hinge point in the two anti-parallel β -strands ('backbone' strands) connecting the two domains. In the domain rotation, helix 5, which is an integral part of domain N2, pivots on helix 11, which remains associated with domain N1. Helix 12 is the connecting peptide to the C-lobe. Disulphide bridges may modify this conformational change in detail; in the C-lobe of lactoferrin (and other transferrins) the C-termini of helices 5 and 11 are joined; in serum transferrin⁷ the C-terminus of helix 5 is joined to the connecting peptide, which also differs structurally from that in lactoferrin⁷.

which makes a third cross-over behind the iron site, and helix 5 (residues 121–136) which is centrally involved in the organization of the N2 domain. These two helices cross at an angle of $\sim 75^\circ$. In the conformational change, helix 5 pivots on helix 11, carrying the N2 domain with it. The rotation axis lies almost exactly along the axis of helix 11. There are slight adjustments in the packing of the two helices, but the rest of the N2 domain seems to be little affected. Superposition of each of the individual domains onto the corresponding domains in iron-lactoferrin shows that the differences in the N2 domain (r.m.s. deviation 0.65 Å for all C_α atoms) are only slightly greater than those for the other three domains (N1, 0.52 Å; C1, 0.43 Å; C2, 0.44 Å).

The opening of the cleft exposes three basic side chains, previously buried within it; Arg 121, Arg 210 and Lys 301. These may serve to attract the CO_3^{2-} anion as the first step in binding. There is, in fact, an area of non-protein density near these side chains; higher resolution refinement will determine whether this represents bound solvent molecules or ions. The position of the hinge is such that the potential iron ligands associate in pairs, Tyr 92 and Tyr 192 with domain N2, Asp 60 and His 253 with domain N1, some 8–10 Å away.

The probable sequence of events in binding suggested by this structure and supported by chemical and kinetic arguments^{14,15}, is that the anion is attracted into the open interdomain cleft and binds first, to the N2 domain at the N-terminus of helix 5 (ref. 6). Four of the six metal ligands (Tyr 92, Tyr 192 and two CO_3^{2-} oxygens) are then in a position to bind the Fe^{3+} to the N2 domain. Binding is completed as N2 rotates and the cleft is closed, Asp 60 and His 253 completing the iron coordination and Asp 60 further linking the two domains by hydrogen bonding⁶.

The closed C-lobe is remarkably similar to that in iron-lactoferrin despite the absence of any metal (Fig. 2); the r.m.s. deviation after superposition is only 0.46 Å and the relative domain movement 0.9° . There does not seem to be any structural reason why both lobes cannot be open at the same time. Modelling the same conformational change on the C-lobe shows that a rotated C2 domain would not clash with any other part of the molecule. In this lobe, however, a disulphide bridge, 483–677, joins the ends of the helices matching 5 and 11 (Fig. 1), creating an extra constraint. A 53° rotation would stretch the distance $C_\alpha^{483}-C_\alpha^{677}$ from 5.2 Å to 8.5 Å, more than the probable maximum for a disulphide bridge¹⁶ (~ 7 Å). Thus, this disulphide probably restricts domain opening in the C lobe, as suggested previously^{4,13,17} but it should not entirely prevent it.

Why, then, is the C-lobe closed? There does not seem to be any metal or anion bound (Fig. 2) and the weakness of coopera-

tive effects in transferrins² is evidence against the opening of the N-lobe having any effect. The most likely explanation is that an equilibrium exists between open and closed forms in solution, as suggested for bacterial periplasmic binding proteins¹⁸, and that the observed structure is selected by crystal packing. Other examples are known where the crystal environment 'freezes' one of several accessible conformations for a flexible protein¹¹. If this is so, the energy difference between open and closed forms must be slight as there are few crystal packing contacts (nine contacts < 3.4 Å for N-lobe atoms, 15 for C-lobe atoms). Furthermore, it implies that if a closed form is relatively stable even in the absence of iron, the additional energy from metal binding must largely account for the great stability of the iron-lactoferrin complex, and of other metal-substituted transferrins.

We note in passing that many solution studies indicate that the C-lobe in transferrins binds more strongly but is kinetically less favoured^{2,17}, and the N-lobe undergoes more facile conformational change¹⁷. As iron release (and presumably cleft opening) is markedly dependent on ionic strength¹⁹, it is possible that under the conditions of crystallization (low ionic strength/alcohol¹⁰) a closed structure for the less flexible C-lobe is a significant solution species. It is also interesting that in melanotransferrin²⁰, it is the C-lobe which seems to have lost the ability to bind iron¹³.

Finally, we have shown previously^{4,13} that each lobe of lactoferrin (and of other transferrins) bears a striking resemblance to the group of bacterial periplasmic binding proteins characterized by Quijcho and co-workers¹⁸. In these proteins, three structural forms have been observed crystallographically: liganded closed, liganded open and unliganded open (although both open and closed structures have yet to be determined for the same protein). In comparing apolactoferrin and iron-lactoferrin, the N-lobe exhibits the same hinge-bending ligand-induced conformational change as do the bacterial proteins, the 'Venus fly-trap' model²¹. The C-lobe of apolactoferrin shows the one remaining form envisaged, closed but unliganded. These observations on lactoferrin thus confirm and extend the principles of structure and function common to this class of binding proteins. □

Received 28 December 1989; accepted 9 February 1990.

1. Aisen, P. & Listowsky, I. *Rev. Biochem.* **49**, 357–393 (1980).
2. Brock, J. H. in *Metalloproteins* Part 2 (ed. Harrison, P. M.) 183–262 (Macmillan, London, 1985).
3. Gorinsky, B. et al. *Nature* **281**, 157–158 (1979).
4. Anderson, B. F. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1769–1773 (1987).
5. Schlabach, M. R. & Bates, G. W. *J. biol. Chem.* **250**, 2182–2188 (1975).
6. Anderson, B. F., Baker, H. M., Norris, G. E., Rice, D. W. & Baker, E. N. *J. molec. Biol.* **209**, 711–734 (1989).
7. Bailey, S. et al. *Biochemistry* **27**, 5804–5812 (1988).

8. Rossenau-Motreff, M. Y. F., Soetewey, R., Lamote, R. & Peeters, H. *Biopolymers* **10**, 1039–1048 (1971).
9. Kilar, F. & Simon, I. *Biophys. J.* **48**, 799–801 (1985).
10. Norris, G. E., Baker, H. M. & Baker, E. N. *J. molec. Biol.* **209**, 329–331 (1989).
11. Bennett, W. S. & Huber, R. *CRC Crit. Rev. Biochem.* **15**, 291–384 (1984).
12. Lesk, A. M. & Chothia, C. *J. molec. Biol.* **174**, 175–191 (1984).
13. Baker, E. N., Rumball, S. V. & Anderson, B. F. *Trends Biochem. Sci.* **12**, 350–353 (1987).
14. Kojima, N. & Bates, G. W. *J. biol. Chem.* **256**, 12034–12039 (1981).
15. Cowart, R. E., Kojima, N. & Bates, G. W. *J. biol. Chem.* **257**, 7560–7565 (1982).
16. Thornton, J. M. *J. molec. Biol.* **151**, 261–287 (1981).
17. Kretschmar, S. A. & Raymond, K. N. *J. Am. Chem. Soc.* **108**, 6212–6218 (1986).
18. Sack, J. S., Saper, M. A. & Quirocho, F. A. *J. molec. Biol.* **206**, 171–191 (1989).
19. Kretschmar, S. A. & Raymond, K. N. *Inorg. Chem.* **27**, 1436–1441 (1988).
20. Brown, J. P. *et al. Nature* **296**, 171–173 (1982).
21. Mao, B., Pear, M. R., McCammon, J. A. & Quirocho, F. A. *J. biol. Chem.* **257**, 1131–1133 (1982).
22. Crowther, R. A. in *The Molecular Replacement Method* (ed. Rossmann, M. G.) 173–178 (Gordon and Breach, New York, 1972).
23. Sussman, J. L. *Meth. Enzymol.* **115**, 271–303 (1985).
24. Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. *Acta crystallogr.* **A43**, 489–501 (1987).
25. Jones, T. A. *J. appl. Crystallogr.* **11**, 268–272 (1978).

ACKNOWLEDGEMENTS. The authors gratefully acknowledge support from the US National Institutes of Health, the Medical Research Council of New Zealand and the New Zealand Dairy Research Institute.

Relaxation of a transfer RNA specificity by removal of modified nucleotides

Véronique Perret, Angela Garcia, Henri Grosjean, Jean-Pierre Ebel, Catherine Florentz & Richard Giegé

Laboratoire de Biochimie, Institut de Biologie Moléculaire et Cellulaire
Centre National de la Recherche Scientifique, 15 rue R. Descartes,
F-67084 Strasbourg Cedex, France

THE molecular recognition of specific transfer RNAs by the appropriate aminoacyl-tRNA synthetase is an important step in determining the accuracy of translation of the genetic message from nucleic acids into proteins. Recent studies using variant tRNAs with specific sequence modifications have indicated particular regions that determine their identity^{1–8}. Here we consider whether the base modifications commonly found in tRNAs⁹ contribute to their identity. Although unmodified tRNA^{Asp} is charged with aspartate as efficiently as the modified native tRNA, it is mischarged with arginine with considerably increased efficiency. Our results indicate that post-transcriptional modification of tRNAs introduces structural 'anti-determinants', restricting the efficiency with which the tRNAs are charged with inappropriate amino acids.

As a model molecule for this study we used yeast tRNA^{Asp}, because its crystal^{10,11} and solution conformations^{12,13} are well known. This tRNA has only eight post-transcriptionally modified nucleotides and, apart from the usual D and Ψ residues, contains only two alkylated bases, m¹G37 in the anticodon loop and m⁵C at position 49 (Fig. 1). Engineering an artificial gene carrying the phage T7 RNA polymerase promoter and expressing it *in vitro* makes it possible to obtain this molecule without modified nucleotides. Because tRNA^{Asp} begins with U1, an unfavourable start for the polymerase¹⁴, a double mutant (G1·C72) was prepared in parallel. This variant and the wild-type transcript are efficiently aspartylated with a catalytic constant (k_{cat}) and a Michaelis constant (K_m) similar to those of native tRNA (Table 1). This indicates that U1 and A72, as well as the eight modified bases, are not identity determinants for tRNA^{Asp}. For U1·A72 this is unexpected, because these residues are uncommon at these positions in yeast tRNAs¹⁵.

Structural studies with chemical probes sensitive to conformational features in RNA^{16,17} pointed to a conformational change of the (G1·C72) transcript as compared with the fully modified molecule (V.P. *et al.*, manuscript in preparation), an observation in agreement with melting and NMR studies on tRNA^{Phe} transcripts^{1,18}. Also significant is the behaviour *in vivo* of tRNA^{Asp} microinjected into *Xenopus* oocytes¹⁹; whereas native tRNA has a half-life of several days, the unmodified (G1·C72) transcript appears more labile after injection. Yet as incubation proceeds this lability is abolished and the stability of the transcript corresponds to that of native tRNA. This property is correlated with the modification *in vivo* of the molecule. Taken together, these results indicate that modified residues have an architectural role in tRNA that affect its structural properties. Furthermore, it is known that mischarging of tRNA is facilitated by organic solvents²⁰, which presumably loosen a compact tRNA structure which would facilitate its interaction with noncognate synthetases²¹. Because tRNA transcripts with an altered conformation could have an increased tendency to be mischarged, we studied the mischarging of tRNA^{Asp} transcripts. As anticipated, tRNA^{Asp}, which is poorly arginylated by arginyl-tRNA synthetase (ArgRS)²², becomes strongly mischarged when unmodified, even with a crude enzyme preparation. This increased arginylation is found in both wild-type and variant transcripts and consequently is insensitive to the replacement in tRNA^{Asp} of the U1·A72 by a G1·C72 base pair (Fig. 2).

The kinetic analysis of the arginylation reactions (Table 1) indicates slightly reduced K_m values for the transcripts as compared to that of modified tRNA^{Asp}, but in contrast shows strongly increased k_{cat} values. Comparison with data obtained

TABLE 1 Kinetic parameters of aspartylation and mischarging of native yeast tRNA^{Asp} and unmodified tRNA^{Asp} by aminoacyl-tRNA synthetases.

tRNA	Aspartyl-tRNA synthetase			Arginyl-tRNA synthetase		
	k_{cat} (s ⁻¹)	K_m (nM)	k_{cat}/K_m	k_{cat} (s ⁻¹)	K_m (nM)	k_{cat}/K_m
tRNA ^{Asp}	0.66	44	100	6.5×10^{-4}	860	7.9×10^{-3}
Unmodified tRNA ^{Asp}						
(G1·C72) variant	0.57	33	115	0.09	390	2.4
(U1·A72) wild type	0.38	28	91	0.045	110	4.3
tRNA ^{Arg/UCU}	—	—	—	0.70	73	100

Pure yeast aminoacyl-tRNA synthetases and native yeast tRNA^{Asp} and tRNA^{Arg/UCU} were obtained by established procedures^{12,22,25,26}. Unmodified yeast tRNA^{Asp} transcripts were synthesized using T7 RNA polymerase: 4 μ g linearized cloned DNA carrying the synthetic yeast tRNA^{Asp} gene (see Fig. 1) were incubated for 3 h at 37 °C in the presence of 2 mM NTP and 8 mM GMP in 40 mM Tris-HCl buffer (pH 8.1) containing 22 mM MgCl₂, 1 mM spermidine, 5 mM dithioerythritol, 40 U RNasin (Promega), 0.55 U pyrophosphatase (Sigma), 0.01% Triton X-100, and 150 U T7 RNA polymerase¹. TLC analysis of an RNase T₂ hydrolysate of the transcripts showed 95% of the molecules start with a 5'-monophosphate. If GMP is omitted, transcripts begin with a 5'-triphosphate; they possess the same aminoacylation properties as the molecules starting with a 5'-monophosphate. The wild-type gene (starting with _{ppp}U1) is transcribed, at best, 90-fold and the variant (_{ppp}G1·C72), 750-fold. Because transcripts are partially heterogeneous in length at the 3'-end, the biologically significant molecules terminating with CCA_{OH} (about 60–70%) were purified by PAGE. Aminoacylation was performed in 50 mM Tris-HCl (pH 7.5), 30 mM KCl, 10 mM ATP, 15 mM MgCl₂, 2.5 mM glutathione, 50 μ M L-[³H]aspartic acid or arginine (2,760 and 1,390 c.p.m. pmol⁻¹, respectively) and the required concentrations of tRNA and synthetase. After incubation at 30 °C, aliquots were spotted on Whatman 3MM filter discs at various times; filters were washed in 5% trichloroacetic acid and ethanol, dried, and their radioactivity counted by liquid scintillation. Transcripts were sensitive to aspartylation conditions: varying the conditions slightly (for example, ionic strength) resulted in almost 50% variation in k_{cat}/K_m , whereas the specificity constant of the modified tRNA remained unchanged.

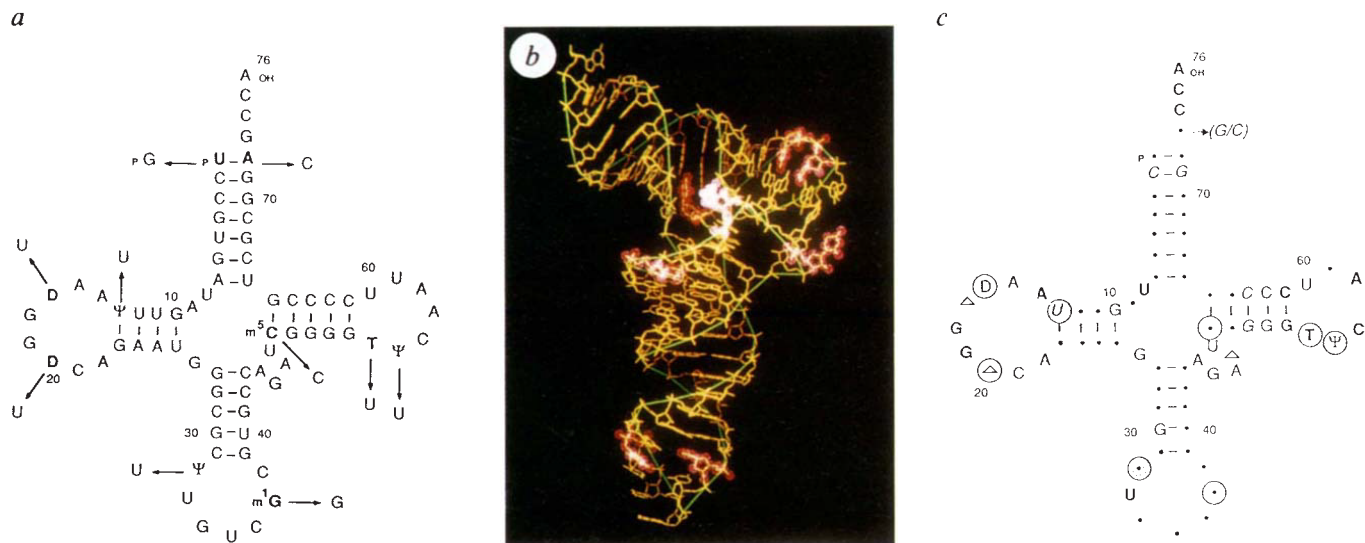


FIG. 1 Transfer RNAs. *a*, Clover-leaf structure of native yeast tRNA^{Asp} (ref. 27) and of the (pG1·C72) variant with positions of modified nucleotides which were replaced by their unmodified counterparts in the T7 RNA polymerase transcripts (substitutions in the transcript are indicated by arrows). *b*, Three-dimensional structure¹¹ of yeast tRNA^{Asp} (without 3'-terminal CCA_{OH}) with positions of modified residues highlighted in pink and Van der Waals' spheres. *c*, Consensus sequence¹⁵ of yeast tRNA^{Asp} and tRNA^{Arg} isoacceptors (invariant positions in elongator tRNAs are shown in bold. Other residues common to aspartic acid and arginine specific tRNAs are in standard lettering, and additional residues common to only yeast tRNA^{Asp} and yeast tRNA^{Arg/UCU}, in italics. Triangles indicate missing residues (note that the D loop and variable region differ in length. Positions modified in yeast tRNA^{Asp} are encircled; the three positions specific to yeast tRNA^{Asp}, Ψ 13, Ψ 32 and m¹G37, are shaded.

METHODS. Ten DNA fragments (17mers to 22mers), corresponding to the yeast tRNA^{Asp} gene sequence with the upstream consensus promoter sequence (positions -17 to -1) of T7 RNA polymerase, were synthesized on an Applied Biosystem 381 A DNA synthesizer and repurified by HPLC on a Nucleosil 120 5c18 column. Oligonucleotides were phosphorylated with T4 polynucleotide kinase in a reaction medium containing 50 μ Ci [γ -³²P]ATP (3,000 Ci mmol⁻¹; Amersham), chased with 50 μ M cold ATP and purified by PAGE on a 15% gel. Hybridization and ligation were performed in 50 mM Tris-HCl (pH 7.5) containing 10 mM MgCl₂, 20 mM dithioerythritol, 0.1 mM spermidine, 1 mM ATP supplemented with plasmid pUC18 digested to XbaI-HindIII sites and 3U T4 DNA ligase (Boehringer). After transfer into *E. coli* JM103, positive clones were isolated by screening plaques for the presence of a supplementary HaeIII restriction site. Sequence of genes and RNA transcripts were verified by DNA and RNA sequencing.

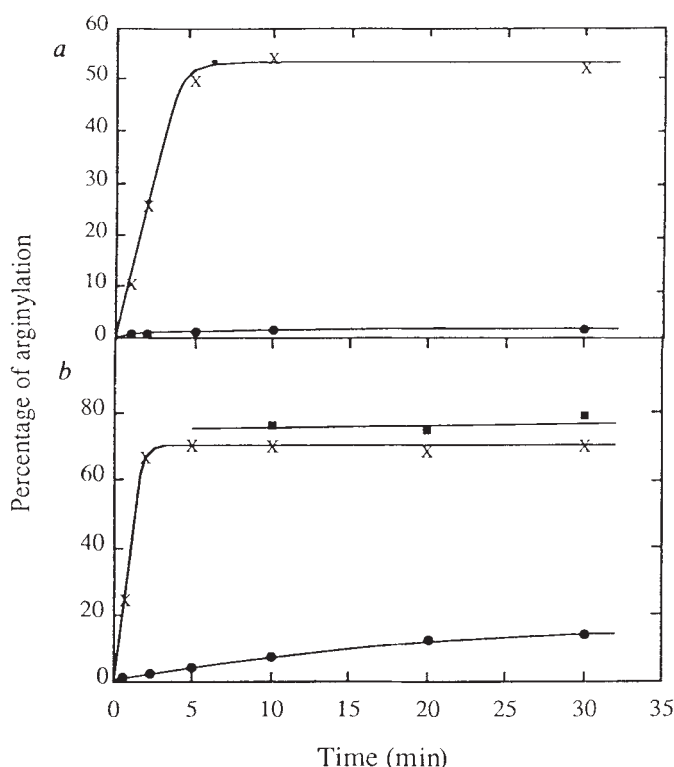


FIG. 2 Arginylation activity of unmodified yeast tRNA^{Asp} transcripts, pG1·C72 variant (x) and pppU1·A72 wild type (■), and of native modified yeast tRNA^{Asp} (●) measured in the presence of a crude enzyme extract from yeast (*a*) or of pure yeast ArgRS (*b*). For aspartic acid-specific tRNAs, mischarging values are normalized to the aspartylation extent of these tRNAs. The mischarging extent of native tRNA^{Asp} is higher with pure ArgRS than with the crude enzyme extract. This is the consequence of the plateau phenomenon which appears when different amounts of synthetase are used in reactions characterized by low k_{cat} (ref. 28). Mischarging of the (pG1·C72) transcript and of native tRNA^{Asp} was assayed with three other synthetases cognate for tRNAs with an A73 discriminator, those specific for histidine, phenylalanine and valine. With these enzymes, no increased mischarging levels of the transcripts were observed.

METHODS. Aminoacylation conditions were as described in Table 1, except for the amino acid that was used, which was L-[¹⁴C]arginine at 580 c.p.m. pmol⁻¹ in *a*. The crude enzyme preparation was a dialysed 100,000g supernatant from bakers' yeast, chromatographed over DEAE-cellulose to remove nucleic acids; 23 μ g total proteins were added to 100 μ l incubation mixture. In *b*, mischarging with ArgRS was performed with an excess of pure enzyme (0.3 μ M).

for tRNA^{Arg/UCU} indicates a 1.5–5-fold decrease in K_m , and thus a slightly better affinity of the cognate tRNA for ArgRS. The k_{cat} values of arginylation of tRNA^{Asp} transcripts are only 8–15-fold less than the k_{cat} of cognate tRNA^{Arg}. Accordingly, in terms of the specificity constant k_{cat}/K_m , tRNA^{Asp} is 12,000-fold less specific than tRNA^{Arg/UCU} for ArgRS, but for unmodified species this decrease is only 20–40-fold. Therefore, specificity of unmodified tRNA^{Asp} for ArgRS becomes 300–500 times greater than that of native tRNA^{Asp}. In other words, unmodified tRNA^{Asp} becomes a quasi-tRNA^{Arg}.

The functional similarity between the aspartic acid- and arginine-specific-tRNAs is not clearly reflected in simple sequence similarities (Fig. 1). How then is discrimination between the two tRNAs achieved? Like other tRNAs⁸, tRNA^{Asp} has identity determinants of which nucleotides from the anticodon region play an important part (A.G. *et al.*, manuscript in preparation). These residues either directly or by allosteric-type effects ensure the tRNA is functionally efficient. But, because of their structural resemblance, tRNAs may mimic each other in their interaction with cognate synthetases. This is the case for both tRNA^{Asp} and tRNA^{Arg/UCU}, which have only a tenfold difference in K_m for arginylation. Thus tRNA^{Asp} interacts with ArgRS and could be mischarged. To reduce this possibility and to minimize the consequences of the incorrect recognition, nature has selected a discrimination mechanism mediated by the post-transcriptional modifications of tRNA^{Asp}. In this mechanism, the modifications act as antideterminants that prevent tRNA^{Asp} becoming a catalytically efficient substrate for ArgRS. This is achieved either by preventing the structural adaptability of tRNA^{Asp} on the false ArgRS (as suggested by the conformational change of the unmodified transcript) and/or by more direct steric effects of the modified nucleotides on interaction sites on the protein. The chemical alterations in tRNA responsible for this antideterminant effect are subtle, as only three modified residues, $\Psi 13$, $\Psi 32$ and m^1G37 , are specific to tRNA^{Asp} when compared with tRNA^{Arg} (Fig. 1). In conclusion, discrimination of tRNA^{Asp} by AspRS and ArgRS is a dichotomic process with positive and negative selections mediated by determinants and antideterminants. Whereas information for positive selection is encoded at the genetic level in DNA, negative selection is dependent on post-transcriptional modifications of tRNA.

This scheme probably applies to other systems and agrees with independent observations, such as the effect of N^2 -methylation of G10 in tRNA on phenylalanylation²³, and that of changing C34 to lysidine in *Escherichia coli* tRNA₂^{Ile} which prevents it from methionylation²⁴. It is, however, premature to conclude that it is the only mechanism ensuring negative selection in tRNA aminoacylation, because for other systems it has been suggested that modified nucleotides do not contribute significantly to tRNA identity⁵. □

20. Giegé, R. *et al.* *Eur. J. Biochem.* **45**, 351–362 (1974).
21. Yarus, M. *Biochemistry* **11**, 2352–2361 (1972).
22. Ebel, J. P. *et al.* *Biochimie* **55**, 547–557 (1973).
23. Roe, B., Michael, M. & Dudock, B. *Nature new Biol.* **246**, 135–138 (1973).
24. Muramatsu, T. *et al.* *Nature* **336**, 179–181 (1988).
25. Lorber, B. *et al.* *Biochem. biophys. Res. Commun.* **117**, 259–267 (1983).
26. Gangloff, J., Schutz, A. & Dirheimer, G. *Eur. J. Biochem.* **65**, 177–182 (1976).
27. Gangloff, J., Keith, G., Ebel, J. P. & Dirheimer, G. *Nature new Biol.* **230**, 125–127 (1971).
28. Dietrich, A., Kern, D., Bonnet, J., Giegé, R. & Ebel, J. P. *Eur. J. Biochem.* **70**, 147–158 (1976).

ACKNOWLEDGEMENTS. We thank F. W. Studier (Brookhaven) for T7 RNA polymerase clone, C. Ehresmann for help during purification of this enzyme, J. Colin and A. Hoefft for oligonucleotide synthesis, E. Westhof for assistance in molecular graphics, P. Remy for use of oocyte microinjection facilities, D. Kern, V. Mikol and M. Ruff for help in synthetase and tRNA preparation, and D. Moras, J. D. Puglisi and J. C. Thierry for discussions. This work was supported by the CNRS, Ministère de la Recherche et de l'Enseignement Supérieur (MRES), NEDO (Japan), and Université Louis Pasteur (Strasbourg).

DNA supercoiling and environmental regulation of virulence gene expression in *Shigella flexneri*

Charles J. Dorman*, Niamh Ni Bhriain* & Christopher F. Higgins†

*Molecular Genetics Laboratory, Department of Biochemistry, Medical Sciences Institute, University of Dundee, Dundee DD1 4HN, UK
†ICRF Laboratories, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

BACTERIAL pathogens undergo profound physiological changes when they infect their hosts, requiring coordinated regulation of gene expression in response to the stresses encountered during infection. A number of environmental stresses (such as osmotic shock and anaerobiosis) have been shown to induce changes in DNA supercoiling that can directly affect the transcription of a specific subset of bacterial genes^{1–6}, at least some of which (the outer-membrane porins and type 1 fimbriae) play a part in bacterial virulence^{7–9}. Here, we demonstrate that the *virR* gene of *S. flexneri*, implicated in the temperature regulation of plasmid-encoded virulence genes^{10–13}, is equivalent to the *osmZ* gene of *Escherichia coli*, which has previously been shown to mediate its regulatory effects through changes in DNA supercoiling. Our results imply that environmentally induced changes in DNA supercoiling are important in the coordinated control of virulence gene expression in *S. flexneri* and have general implications for the control of bacterial virulence.

Genetic complementation has shown that the region of the *E. coli* chromosome containing the *osmZ* locus can partially complement *virR* mutations in *S. flexneri*¹⁴. We therefore investigated whether *virR* and *osmZ* are allelic. *S. flexneri* strain BS184 carries a fusion of *lacZ* to a temperature-regulated invasion gene. We found that expression of this *vir-lac* fusion gene was not only regulated by temperature¹³, but was also dependent on growth phase (Fig. 1a). Significantly, DNA supercoiling varies with growth phase¹.

Strain BS185 is congenic with BS184, except that it carries a Tn10 insertion in the *virR* gene¹³. The *virR::Tn10* insertion was transduced, through a restriction-deficient intermediate, into a strain of *E. coli* carrying a *lacZ* fusion to *proU*, an osmotically induced operon whose expression is derepressed by *osmZ* mutations². Transductants (for example, CJD389) showed derepression of *proU* expression that is typical of *osmZ* mutants (Table 1). These transductants were also derepressed for expression of the *bgl* operon (Table 1) and showed a mucoid phenotype, which are both characteristics of *osmZ* mutants². Southern blotting showed that the *virR::Tn10* had not transposed but had inserted into the *E. coli* chromosome at a site corresponding to that of the *virR::Tn10* insertion in *S. flexneri*

Received 2 January; accepted 27 February 1990.

1. Sampson, J. & Uhlenbeck, O. C. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1033–1037 (1988).
2. Hou, Y.-H. & Schimmel, P. *Nature* **333**, 140–145 (1988).
3. McClain, W. H. & Foss, K. *Science* **240**, 793–796 (1988).
4. Rogers, J. & Söll, D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6627–6631 (1988).
5. Schulman, L. H. & Pelka, H. *Science* **242**, 765–768 (1988).
6. Sampson, J. R., DiRenzo, A. B., Behlen, L. S. & Uhlenbeck, O. C. *Science* **243**, 1363–1366 (1989).
7. Schimmel, P. *Biochemistry* **28**, 2747–2759 (1989).
8. Normanly, J. & Abelson, J. A. *Rev. Biochem.* **58**, 1029–1049 (1989).
9. Björk, G. R. *et al.* *A. Rev. Biochem.* **56**, 263–287 (1987).
10. Moras, D. *et al.* *Nature* **288**, 669–674 (1980).
11. Westhof, E., Dumas, P. & Moras, D. *J. molec. Biol.* **184**, 119–145 (1985).
12. Romby, P. *et al.* *J. molec. Biol.* **184**, 455–471 (1985).
13. Romby, P., Moras, D., Dumas, P., Ebel, J. P. & Giegé, R. *J. molec. Biol.* **195**, 193–204 (1987).
14. Dunn, J. J. & Studier, F. W. *J. molec. Biol.* **166**, 477–535 (1983).
15. Sprinzl, M., Hartmann, T., Weber, J., Blank, J. & Zeidler, R. *Nucleic Acids Res.* **17**, 1–172 (1989).
16. Peattie, D. A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4679–4682 (1980).
17. Krzyzosiak, W. J. *et al.* *Biochemistry* **27**, 5771–5777 (1988).
18. Hall, K. B., Sampson, J. R., Uhlenbeck, O. C. & Redfield, A. G. *Biochemistry* **28**, 5794–5800 (1989).
19. Grosjean, H. & Kubli, E. in *Microinjection and Organelle Transplantation Techniques* (eds Celis, J. E., Graessman, A. & Loyer, A.) 304–326 (Academic, New York, 1986).

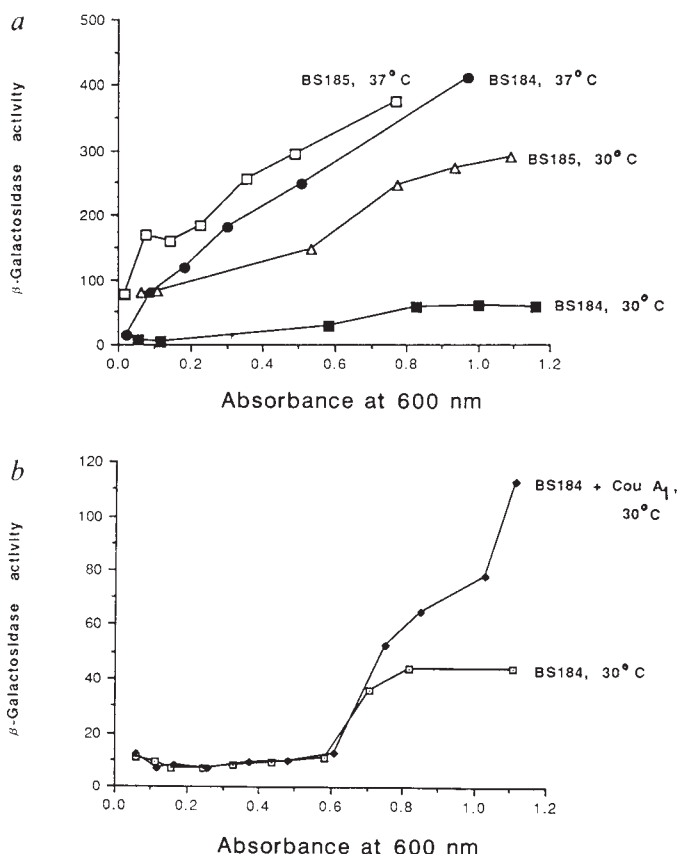
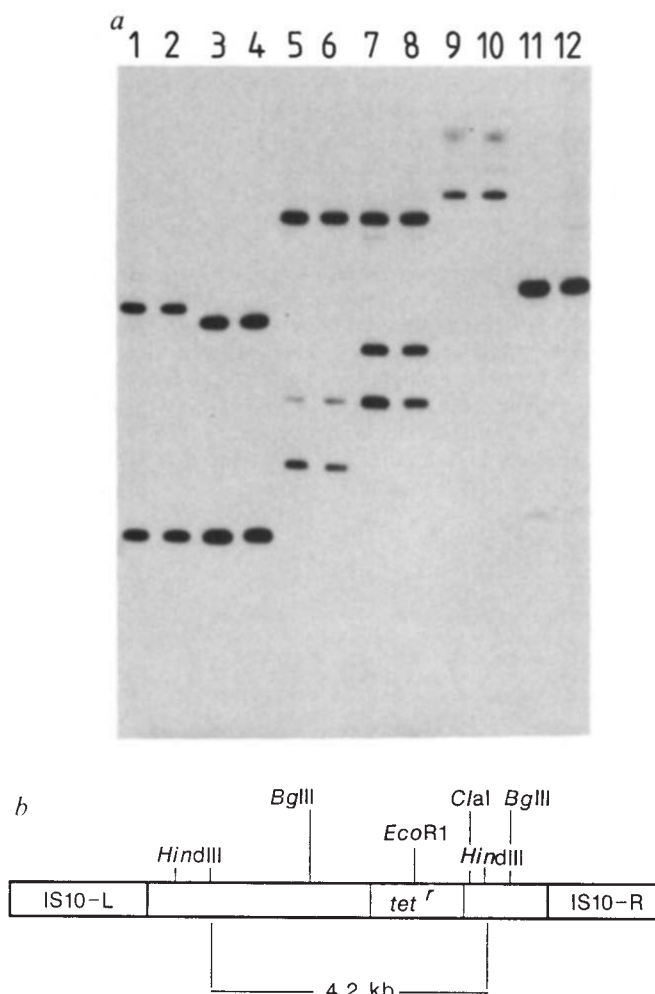


FIG. 1 Temperature and growth phase regulation of the *vir-83::Mud1734* transcriptional fusion in *S. flexneri*. *S. flexneri* strain BS184 carries a transcriptional fusion (*vir-83::Mud1734*) between *lacZ* (the structural gene for β -galactosidase) and a virulence gene on its 220-kilobase (kb) virulence plasmid¹³. Expression of *lacZ* is under the control of a temperature-regulated virulence gene promoter. This promoter is repressed at 30 °C, induced at 37 °C and its activity can be monitored by measuring β -galactosidase activity¹². Strain BS185 is congenic with BS184 but carries a *virR::Tn10* mutation. In this strain, expression of the *vir-lac* fusion is derepressed at 30 °C and this derepression correlates with the derepression of several other immunologically detectable virulence factors¹³. **a**, Effect of growth phase and temperature on β -galactosidase activity in BS184 and BS185. **b**, Effect of coumermycin A₁ (20 μ g ml⁻¹) on β -galactosidase activity in BS184 grown at 30 °C. Previous work has shown that the concentrations of gyrase inhibitors used here relax plasmid DNA¹. All strains were grown as 35 ml trypticase soy broth cultures in 500-ml flasks with vigorous aeration. β -Galactosidase activity was assayed as described¹⁸. Data shown are the means of at least four experiments. Standard errors were less than 5%.

FIG. 2 Analysis of the bacterial chromosome in the neighbourhood of the *Tn10* insertion associated with *virR*. **a**, Southern blot of chromosomal DNA from *S. flexneri* strains BS185 (lanes 1, 5, 9) and CJD403 (lanes 2, 6, 10) and from *E. coli* strains CJD389 (lanes 3, 7, 11) and CJD390 (lanes 4, 8, 12). BS185 is a *virR::Tn10* derivative of BS184 (ref. 13). CJD403 is a phenotypic revertant of BS185, described in the text. CJD389 is a *virR::Tn10* derivative of *E. coli* strain GM37 (*proU-lacZ*)². CJD390 is a *virR::Tn10* derivative of GM37 which is phenotypically *virR*⁺. DNA was digested with *Bgl*III (lanes 1–4); *Eco*RI (lanes 5–8); *Clal* (lanes 9–12). The probe used was a 4.2-kb *Hind*III fragment from *Tn10*, isolated from plasmid pNK81 (ref. 17). The DNA was radiolabelled using random oligonucleotide priming¹⁹ and hybridized to chromosomal DNA, which has been transferred to a Hi-Bond N filter using standard techniques²⁰. Within each pair of tracks, no differences in hybridization pattern were detected. Differences between species reflect variations in nucleotide sequence between *S. flexneri* and *E. coli*. A common *Eco*RI fragment of about 16 kb is present in all four strains. Presumably, this reflects the high level of DNA sequence homology reported for these two species²¹. The common band seen in the *Bgl*III digests is an internal *Tn10* sequence. Additional bands seen in the *Eco*RI digests may be due to a rearrangement involving the *tetA* gene of *Tn10*. The unique *Eco*RI site in *Tn10* occurs within this gene and it may have been duplicated within *S. flexneri* and the altered transposon transduced to *E. coli*. **b**, Diagram of *Tn10* showing the source of the probe DNA and the locations of relevant restriction sites²².



(Fig. 2a). Changes in linking number of reporter plasmid DNA are indicative of changes in the *in vivo* level of DNA supercoiling in bacterial cells^{1-3,15}. Importantly, analysis of reporter plasmid topoisomer distributions in *virR*::Tn10 derivatives of *E. coli* (for example, CJD389) also showed changes in linking number similar to those observed for *osmZ* mutants (Fig. 3c).

These data show that *virR* and *osmZ* are allelic, implicating changes in DNA supercoiling in the control of *vir* gene expression. To confirm that changes in DNA supercoiling play a part in temperature regulation of gene expression in *Shigella*, we analysed topoisomer distributions in *S. flexneri* strain BS184 and its *virR*::Tn10 derivative BS185, using the small, endogenous plasmids as reporters. The phenotypic effect of the *virR*::Tn10 mutation in *S. flexneri* strain BS185 mimics the effect of growth of BS184 at 37 °C: both cause derepression of the *vir-lac* fusion gene. If DNA supercoiling affects expression of *vir-lac*, then plasmid supercoiling in BS184 grown at 37 °C should be altered in the same way as that seen in BS185 at 30 °C. This is indeed the case (Fig. 3d); there is a good correlation between *vir-lac* expression and DNA supercoiling, whether affected by temperature or by the *virR*::Tn10 insertion. Unexpectedly, plasmid DNA from BS185 grown at 30 °C was more relaxed than that from BS184 at 30 °C, with a difference in linking number of 3 (Fig. 3a). Thus, the effect of the *virR*::Tn10 mutation on plasmid DNA supercoiling in *S. flexneri* is opposite to that of the same *virR*::Tn10 mutation (or of *osmZ* mutations) in *E. coli* (Fig. 3). Similarly, the effect of temperature on plasmid DNA supercoiling in *E. coli* was opposite to that found for *S. flexneri* (Fig. 3d)⁶. *Salmonella typhimurium*, another facultative intracellular pathogen, behaved like *S. flexneri* (Fig. 3d).

The reason why the *osmZ(virR)* mutations and temperature affect plasmid DNA supercoiling differently in *E. coli* and *S. flexneri* is not understood, although it is probably a function of the complexity of the *osmZ(virR)* locus.

Phenotypic revertants of the BS185 *virR*::Tn10 mutation arose at a high frequency (5×10^{-2}) and we decided to analyse one of them (CJD403) further. The *vir-lacZ* gene fusion was shown, by marker rescue, to be unaltered in CJD403, and transduction experiments showed that the mutation responsible for the phenotypic reversion is closely linked (100%; $n = 50$) to the Tn10 insertion. Southern blotting showed that the *virR*::Tn10 in CJD403 had not transposed to a new site and revealed no obvious differences in chromosome structure close to the *virR*::Tn10 (Fig. 2). Importantly, plasmid DNA was found to be more highly supercoiled in CJD403 grown at 30 °C than it was in its immediate parent, BS185 (Fig. 3b). Thus, the mutation responsible for the phenotypic reversion of CJD403 also increases plasmid supercoiling and these supercoiling changes are indeed responsible for changes in *vir* gene expression.

If *virR* in *S. flexneri* is equivalent to *osmZ* of *E. coli*, then DNA supercoiling changes mediated by other means would also be expected to regulate *vir-lacZ* expression. To investigate whether the *vir-lacZ* promoter is sensitive to changes in DNA supercoiling, we used coumermycin A₁ and novobiocin, drugs that inhibit DNA gyrase¹⁴. Both gyrase inhibitors induced *vir-lacZ* expression in BS184, although not as much as the combined effects of temperature and growth phase (Fig. 1b).

Our results are strong evidence that environmentally induced changes in DNA supercoiling are important in the regulation

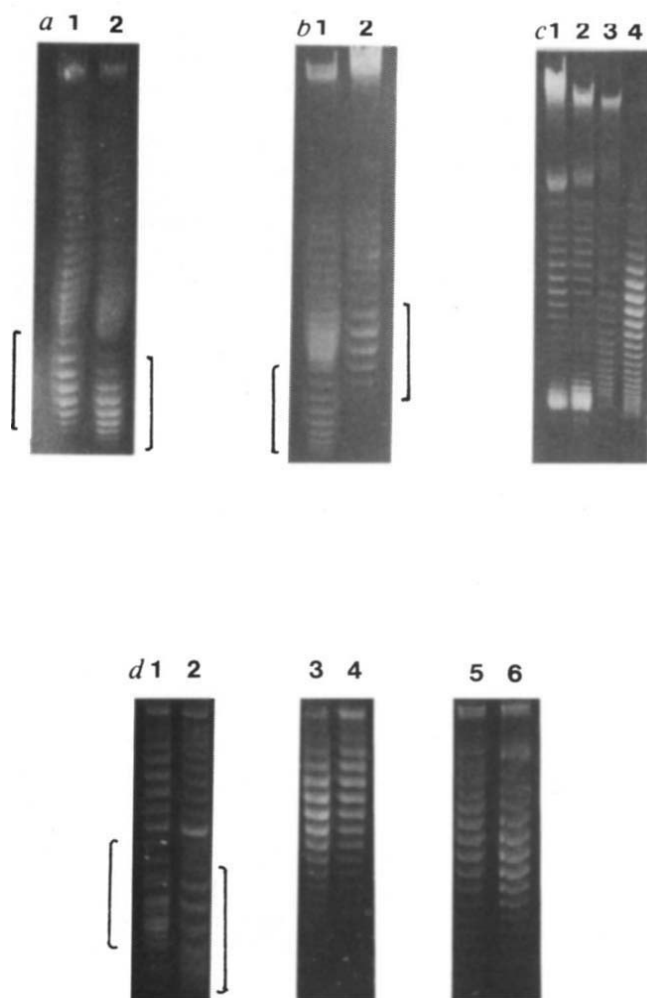


FIG. 3 Distribution of reporter plasmid DNA topoisomers. a, Distribution of topoisomers of the 2-kb and 4-kb indigenous *S. flexneri* plasmids isolated from BS184 (*virR*⁺, *vir-lacZ*) (lane 1) and BS185 (*virR*::Tn10, *vir-lacZ*) (lane 2) (overnight cultures grown in trypticase soy broth at 30 °C). b, Distribution of topoisomers of the indigenous *S. flexneri* plasmids isolated from BS185 (30 °C, lane 1) and its phenotypic revertant, CJD403 (30 °C, lane 2). c, Distribution of topoisomers of plasmid pACYC184 isolated from *E. coli* strains grown at 37 °C: MC4100 (lane 1), GM37 (*proU-lacZ*; lane 2), CJD389 (*virR*::Tn10 derivative of GM37; lane 3) and BRE2076 (an *osmZ* derivative of GM37; lane 4). d, Distribution of topoisomers of the indigenous *S. flexneri* plasmids isolated from BS185 grown at 30 °C (lane 1) and 37 °C (lane 2) (overnight cultures grown in trypticase soy broth); distribution of topoisomers of pACYC184 plasmid DNA isolated from *E. coli* strain MC4100 grown at 30 °C (lane 3) and 37 °C (lane 4)—similar results have been reported previously¹⁵; distribution of topoisomers of pACYC184 plasmid DNA isolated from *S. typhimurium* strain LT-2 grown at 30 °C (lane 5) and 37 °C (lane 6). The *E. coli* and *S. typhimurium* strains used in this study have been described². Plasmid DNA was extracted from *S. flexneri* and *E. coli* by a standard alkaline lysis procedure²³. Plasmid DNA was electrophoresed on 1% agarose gel (BRL) using standard Tris-borate-EDTA (TBE) buffer²¹. Gels contained 25 µg ml⁻¹ chloroquine (Sigma). Under these conditions, those molecules most relaxed before extraction from the bacteria migrate most rapidly through the gel. Gels were soaked in water for 6 h to remove the chloroquine, then stained with ethidium bromide (5 µg ml⁻¹) and photographed. The brackets flanking the gels show the extent of the topoisomer distribution for the 2-kb *S. flexneri* indigenous plasmid, where appropriate.

TABLE 1 Effect of *virR::Tn10* on *proU-lacZ* and *bgl* expression in *E. coli*

Strain*	β -Galactosidase activity	
	No NaCl	0.3 M NaCl
GM37	8	1,336
BRE2076	166	Dead†
CJD389	423	2,316
CJD390	7	1,270

Strain*	β -Glucosidase activity	
	No NaCl	0.3 M NaCl
GM37	10	
BRE2076	301	
CJD389	233	
CJD390	8	

* Genotypes of strains are given in the legends to Figs 2 and 3.

† Osmotically sensitive².

β -Galactosidase assays for *proU-lacZ* expression were performed on SDS-chloroform permeabilized cells outgrown in 2 ml trypticase soy broth cultures in test tubes to mid-log (with or without 0.3 M NaCl) as described^{2,18}. β -Glucosidase assays for *bgl* expression were on cells grown in MMA (minimal medium A) + 0.4 M succinate in the presence of 5 mM β -methyl-D-glucoside, a gratuitous inducer of *bgl*. Cells were assayed² for their ability to transport, phosphorylate and metabolize the β -glucoside derivative, *o*-nitrophenyl- β -D-pyranoglucoside. MMA + 0.4 M succinate was a poor growth medium for CJD389. This is also an unexplained feature of several *osmZ* alleles in *E. coli* (unpublished data). Values are averages of three independent experiments. Standard errors were less than 10%.

of *vir* gene expression in *S. flexneri*. Furthermore, we have shown that there is a close genetic relationship between *virR* and *osmZ*, a gene known to be important in the environmental control of DNA supercoiling. As we predicted^{2,16}, one of the products of the complex *osmZ(virR)* locus is a small, histone-like protein (known as H1 or HNS) (C. J. Hulton, A. Seirafi and C.F.H., manuscript in preparation). We have argued elsewhere^{7,16} that environmentally determined levels of DNA supercoiling are important in regulating gene expression and have suggested that such a mechanism could be exploited by bacterial pathogens to adapt to new environments encountered at each stage of the infection process. Environmentally induced changes in DNA supercoiling must act with specific regulators of gene expression (such as the well characterized family of two-component regulators¹⁷). We have demonstrated that changes in DNA supercoiling in response to environmental factors do indeed contribute to the control of bacterial virulence. □

16. Higgins, C. F., Dorman, C. J. & Ni Bhriain, N. in *The Bacterial Chromosome* (ed Drica, K. & Riley, M.) 421-432 (ASM, Washington DC, 1990).
17. Gross, R., Arico, B. & Rappuoli, R. *Molec. Microbiol.* **3**, 1661-1667 (1989).
18. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1972).
19. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1984).
20. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
21. Brenner, D. J., Fanning, G. R., Sherman, F. J. & Falkow, S. *J. Bact.* **109**, 953-965 (1972).
22. Foster, T. J., Davis, M. A., Roberts, D. E., Takeshita, K. & Kleckner, N. *Cell* **23**, 201-213 (1981).
23. Ish-Horowitz, D. & Burke, J. F. *Nucleic Acids Res.* **9**, 2989-2994 (1981).

ACKNOWLEDGEMENTS. We thank P. J. Sansonetti for discussion and A. T. Maurelli for bacterial strains. C.J.D. is a Royal Society University Research Fellow. C.F.H. is supported by the ICRF. This work was supported by the Wellcome Trust and the Society for General Microbiology.

CORRECTION

A new mechanism for induced vitamin D deficiency in calcium deprivation

M. R. Clements, L. Johnson & D. R. Fraser

Nature **325**, 62-65 (1987).

FURTHER work by the authors has revealed that the above paper contained a systematic miscalculation in the determination of plasma elimination half times for 25-hydroxyvitamin D in rats. The values quoted in the paper (mainly in Fig. 1 and Table 1) are over-estimates by a factor of log 10*e*. To obtain the correct values the published figures need to be divided by a factor of 2.3. This correction does not alter the validity or the interpretation of the published observations. Indeed, because the half time for 25-hydroxyvitamin D in the circulation is significantly shorter than previously indicated, the detrimental effect of calcium deprivation on vitamin D status is even greater than the authors had suggested. □

ERRATUM

DNA cleavage catalysed by the ribozyme from *Tetrahymena*

Daniel Herschlag & Thomas R. Cech

Nature **344**, 405-409 (1990).

IN Figs 1 and 2 of the above article, the shaded regions referred to in the figure legend did not show up because of an oversight in the reproduction process. The 5' sequence (GGGAGG-5') of the ribozyme (shown seven times) should have been shaded in these figures. Also, the complex formed between the ribozyme (E) and its oligonucleotide product (P) was changed to E~P in the editing process. Because this bond symbol might wrongly implicate participation of a high-energy phosphate bond in complex formation, this complex should have been represented by E·P. Other complexes involving the ribozyme, its oligonucleotide substrate (S), and G should then have been written as E·G·S, or E·G. □

Received 21 December 1989; accepted 22 February 1990.

1. Dorman, C. J., Barr, G. C., Ni Bhriain, N. & Higgins, C. F. *J. Bact.* **170**, 2816-2862 (1988).
2. Higgins, C. F. *et al. Cell* **52**, 569-584 (1988).
3. Ni Bhriain, N., Dorman, C. J. & Higgins, C. F. *Molec. Microbiol.* **3**, 933-942 (1989).
4. Graeme-Cook, K. A., May, G., Bremer, E. & Higgins, C. F. *Molec. Microbiol.* **3**, 1287-1294 (1989).
5. Balke, V. L. & Gralla, J. D. *J. Bact.* **169**, 4499-4504 (1987).
6. Goldstein, E. & Drica, K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4046-4050 (1984).
7. Higgins, C. F. & Dorman, C. J. in *Bacterial Protein Toxins* (ed. Rappuoli, R. *et al.*) 293-302 (Gustav and Fisher, Stuttgart and New York, 1990).
8. Dorman, C. J., Chatfield, S., Higgins, C. F., Hayward, C. & Dougan, G. *Infect. Immun.* **57**, 2136-2140 (1989).
9. Dorman, C. J. & Higgins, C. F. *J. Bact.* **169**, 3840-3843 (1987).
10. Sansonetti, P. J., Kopecko, D. J. & Formal, S. B. *Infect. Immun.* **35**, 852-860 (1982).
11. Yoshikawa, M. *et al. Microbiol. Sci.* **5**, 333-339 (1988).
12. Maurelli, A. T., Blackmon, B. & Curtiss III, R. *Infect. Immun.* **34**, 195-201 (1984).
13. Maurelli, A. T. & Sansonetti, P. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2820-2824 (1988).
14. Hromockyj, A. E. & Maurelli, A. T. *J. Bact.* **171**, 2879-2881 (1989).
15. Drica, K. *Microbiol. Rev.* **48**, 273-289 (1984).

DNA mutagenesis and recombination

D. H. Jones, K. Sakamoto, R. L. Vorce and B. H. Howard

The polymerase chain reaction is used for site-specific mutagenesis and for DNA recombination without any enzymatic reaction *in vitro*, apart from DNA amplification.

The polymerase chain reaction (PCR) has become a rapid method for site-specific mutagenesis¹⁻⁷ and for DNA recombination^{8,9}. As PCR amplification is not designed to generate cohesive ends, the cloning of PCR products has required additional *in vitro* enzymatic manipulations following the PCR amplification.

We recently described a novel method in which the formation of DNA joints is achieved by using separate PCR amplifications to generate products that when combined, denatured and reannealed form double-stranded DNA with discrete, cohesive single-stranded ends¹⁰. As these single-stranded ends are designed to anneal to each other to yield circles, the approach is termed recombinant circle PCR (RCPCR). RCPCR-generated DNA circles form without restriction enzyme digestion or ligation, and can be transfected directly into *Escherichia coli*.

Site-specific mutagenesis

Point mutagenesis by RCPCR is shown in Fig. 1. The plasmid is mutated in two separate PCR amplifications. In each amplification, the identical base pair is mutated by incorporating a primer with a base mismatch into the PCR product (primer 1 in one amplification and primer 3 in the other). However, the positioning of the primers makes the 'break point' of

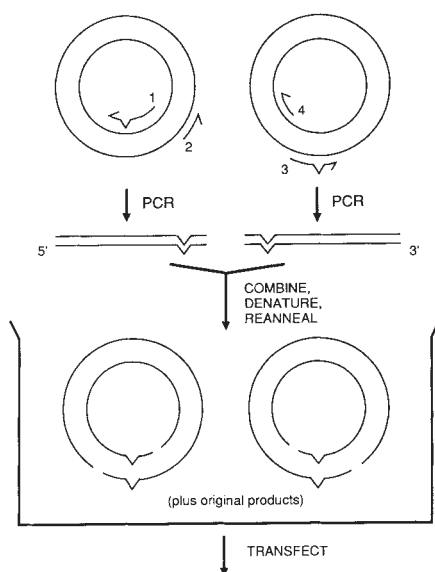


FIG. 1 Generation of a point mutation by recombinant circle PCR (RCPCR). Numbered hemiarrows, primers; notches, designate point mismatches in the primers and resulting mutations in the PCR products.

the plasmid in one amplification distinct from the 'break point' of the plasmid in the other amplification. This 'break point' is the region between the two 5' ends of the amplifying primers in any one PCR amplification. Because of the differing positions of these break points, when a strand from one PCR product anneals to its complementary strand from the other PCR product, a discrete single-stranded segment extends from each end. As these single strands are complementary to each other, they anneal, forming DNA circles that can be transfected into *E. coli* and repaired *in vivo*.

RCPCR-mediated site-specific mutagenesis has been accomplished in constructs up to 3.1 kb using 14 amplification cycles, and in a construct of 6.1 kb using 18 amplification cycles. Protocols have included point mutagenesis or the insertion of a 35-bp segment (Fig. 2). Insertion

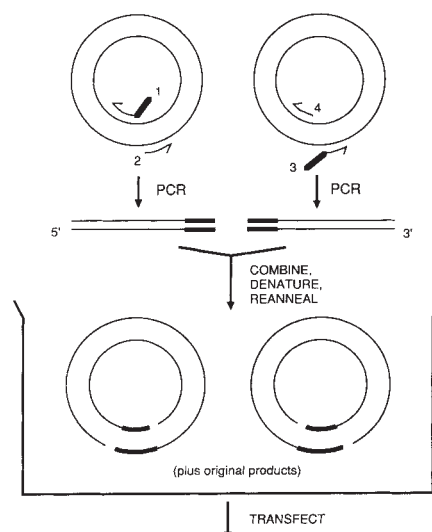


FIG. 2 Insertional mutagenesis by RCPCR. The heavy lined 5' regions of primers 1 and 3 are the complementary strands of the segment that is to be inserted.

of 85 bp has also been achieved, by setting the break points within the region to be inserted (P. Gershon, personal communication). In each case, the two PCR products were purified by agarose gel electrophoresis and then combined, denatured and reannealed before transfection into *E. coli*. Using an extension time at 72 °C of one minute per kilobase during each amplification cycle, 83 per cent or more of the clones contained the mutation of interest. Furthermore, successful mutagenesis following generation of

gapped circles, in addition to nicked circles, has broadened the choice of primer sites and has lessened the cost of making multiple mutations by allowing the use of conserved 'outside' primers (primers 2 and 4), while using pairs of mutating primers that span different areas. Small gaps (up to 19 bp tested) in the recombinant circles have not diminished the transfection efficiency or the yield of mutants.

The critical parameter in determining success with this method has been careful design of the primers. In particular, the 3' portion (11 bp) of each primer is examined for unwanted annealing sites to the construct. In the point-mutagenesis protocol, if the mismatched nucleotide is located towards the 3' end of each mutating primer (primers 1 and 3), the strand overlap comprising the joint of each recombinant circle is correspondingly increased. Generally, when designing primers that mutate one or several base pairs, we have positioned 20 nucleotides 5' to the mismatched nucleotide(s), and 6–10 nucleotides 3' to the mutagenesis region. Therefore, in a point-mutagenesis protocol, the strand overlap in the joint that holds together the recombinant circle is 41 bp (20 bp + the substituted bp + 20 bp).

DNA recombination

RCPCR can also be used to recombine DNA. Figure 3 illustrates a strategy for moving a segment of one plasmid into another plasmid at a specific location and with a defined orientation, using no

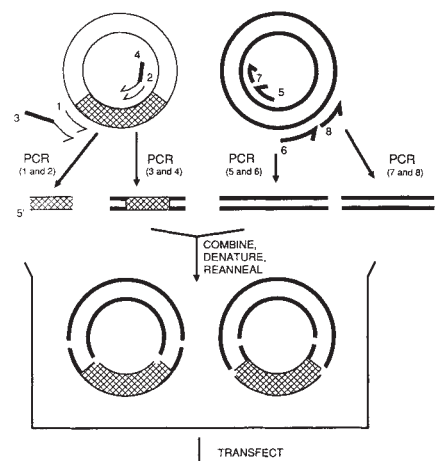


FIG. 3 DNA recombination by RCPCR. Thin circles, DNA strands of the donor plasmid; thick circles, DNA strands of the recipient plasmid; cross-hatched area, insert segment.

enzymatic reaction apart from PCR. The segment that is to be inserted into the recipient plasmid is amplified using primers 1 and 2. In a separate reaction, this insert is both amplified and modified using primers 3 and 4. The 5' regions of primers 3 and 4 are homologous to the recipient plasmid, and these sequences are incorporated into the PCR product. Therefore, when one reanneals the products of these two PCR amplifications, DNA arises with single-stranded ends that can anneal to recipient plasmid sequences. The recipient plasmid is amplified with primers 5 and 6. Primers 5 and 6 are complementary to the 5' regions of primers 3 and 4 (or 4 and 3, depending on the orientation of the insert desired in the RCPCR product).

In a separate reaction, the recipient plasmid is amplified with primers 7 and 8. The 5' ends of primers 7 and 8 are internal to the 5' ends of primers 5 and 6 in the PCR products. Therefore, when the product from primers 5 and 6, and the product from primers 7 and 8, are combined, denatured, and reannealed, DNA is formed with single-stranded ends that will anneal to the single-stranded sequences protruding from the insert sequence. Note that as long as the 5' regions of primers 3 and 4 are complementary to the 5' regions of primers 5 and 6, these four regions do not have to have homology to either donor or recipient plasmid sequences. The products from all four PCR amplifications are purified, and can be combined, denatured and reannealed in a single mixture.

In a recombination protocol, mis-oriented and/or mispositioned primers are likely errors. In general, it is easiest to visualize the recombinant circles by drawing out the sequence comprising each joint. A test of this strategy, with a 30-bp region of overlap in each joint, yielded clones with inserts in the recipient plasmid at an efficiency of 30 per cent¹⁰.

Advantages and limitations

Previous reports pioneered recombinant PCR by using a process called splicing by overlap extension for site-specific mutagenesis and for DNA recombination^{1,4,8}. RCPCR has several advantages over this earlier approach. Splicing by overlap extension requires two sequential PCR amplifications and generates a product that must undergo a conventional cloning step into a vector. RCPCR permits mutagenesis and recombination with only one set of PCR amplifications, and simultaneously accomplishes the cloning step. The recombinant DNA circles are transfected into *E. coli* without phosphorylation of the primers or products, and without treatment with Klenow fragment, restriction enzymes, or ligase.

Polymerase errors are a concern in any protocol that uses PCR-generated prod-

The molecular marketplace

What follows is a selection of what's new in molecular biology and features kits for the non-isotopic labelling of oligonucleotides and the purification of plasmid DNA from *E. coli*, and Mac-based sequence software.

UNITED STATES Biochemical has introduced three new kits — the Base Δ , Codon Δ and Multi-Base F **excision linker kits** — designed to create precise nucleotide changes in cloned genes when used in conjunction with class IIS restriction endonucleases (*Reader Service No. 101*). Unlike site-directed mutagenesis, USB says excision linker mutagenesis can be used to generate multi-site as well as single-site nucleotide changes, including codon deletion (with Codon Δ , \$175 (US)) and nucleotide substitutions (with Base Δ , \$200 (US)). The \$175 (US) Multi-Base F kit is designed to generate multiple base deletions of either 18 or 26 nucleotides.

For protein blotting/sequencing applications, Schleicher and Schuell recommends the use of its Westran hydrophobic PVDF **membrane** (*Reader Service No. 102*). Westran PVDF membrane can be used for Western blotting of proteins and has a high affinity for low molecular weight proteins. The membrane can also



S&S's Westran PVDF membranes.

be used as a support for protein sequencing by Edman degradation. The electro-transferred or dot-blotted proteins remain tightly bound during sequencing, producing high and reproducible yields of amino acids in Edman digestion, says S & S.

As a complement to its range of designer probes, British Bio-technology is offering two new **kits for the labelling of oligonucleotide probes** at the 3' and 5' ends (*Reader Service No. 103*). The kits are designed for radiolabelling probes with a high specific activity for use in Southern

ucts¹¹. As polymerase errors are cumulative during PCR, the low number of cycles necessary to generate the recombinant of interest with RCPCR should reduce the risk of acquiring a base error¹¹. Because of the possibility of errors in single clones following PCR amplification, researchers may want to test more than one product in a functional assay, or may choose to reclone a restriction fragment that contains the mutated or recombined fragment into a plasmid that has not undergone PCR amplification.

Applications of RCPCR

Numerous applications for the RCPCR method can be envisaged. By modification of the primers incorporated into the ends of the PCR products, mutagenesis by RCPCR could be used to generate deletions² as well as for point mutagenesis and insertional mutagenesis. Also, mutagenesis could be combined with DNA recombination. DNA recombination by RCPCR makes it possible to transfer one DNA molecule into any point of another DNA molecule, with or without replacing a segment in the recipient DNA. Therefore, RCPCR provides a rapid method for gene fusion, or for changing the orientation of a DNA segment (for example, promoter element). RCPCR permits rapid and precise vector reconstruction

without regard to enzyme restriction sites, and without the use of linkers or adaptors. The time from setting up the PCR amplification to transfection into *E. coli* is 24 hours. Furthermore, phosphorylation of the primers at the 5' terminus followed by ligation of RCPCR products could be used to generate closed circular DNA *in vitro*. In summary, use of RCPCR facilitates the rapid generation of site-specific mutants and DNA recombinants. □

Douglas H. Jones, Kazuichi Sakamoto, Roseann L. Vorce and Bruce H. Howard, Laboratory of Molecular Biology, Building 37, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. For more information fill in Reader Service Number 100.

1. Higuchi, R., Krummel, B. & Saiki, R.K. *Nucleic Acids Res.* **16**, 7351–7367 (1988).
2. Vallette, F., Mege, E., Reiss, A. & Adesnik, M. *Nucleic Acids Res.* **17**, 723–733 (1989).
3. Kadowaki, H., Kadowaki, T., Wordisford, F.E. & Taylor, S.I. *Gene* **76**, 161–166 (1989).
4. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. & Pease, L.R. *Gene* **77**, 51–59 (1989).
5. Nelson, R.M. & Long, G.L. *Analyt. Biochem.* **180**, 147–151 (1989).
6. Kammann, M., Laufs, J., Schell, J. & Gronenborn, B. *Nucleic Acids Res.* **17**, 5404 (1989).
7. Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G. & Galas, D.J. *Nucleic Acids Res.* **17**, 6545–6551 (1989).
8. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. & Pease, L.R. *Gene* **77**, 61–68 (1989).
9. Yon, J. & Fried, M. *Nucleic Acids Res.* **17**, 4895 (1989).
10. Jones, D.H. & Howard, B.H. *BioTechniques* **8**, 178–183 (1990).
11. Saiki, R.K. *et al. Science* **239**, 487–491 (1988).

and Northern, colony and plaque, *in situ*, and dot-blot hybridizations. Designer probes have been labelled with a specific activity of 10^9 d.p.m. μg^{-1} using the 5' kit (£55 (UK), 20 reaction (£55 (UK), 20 reaction) and 10^{10} d.p.m. μg^{-1} using the 3' kit (£95 (UK), 15 reaction).

Epicentre Technologies is offering Ampligase, a **thermostable DNA ligase** which is active over a 15-85 °C temperature range, in kit form (Reader Service No. 104). It is most suitable for use in nick ligation assays, for synthesis of genes using overlapping oligodeoxynucleotides, or for other manipulations of DNA that involve ligation of nicks in duplex DNA. Under appropriate assay conditions, even a single base mismatch with the template should be detected, says Epicentre. Although ligation assays with non-thermostable ligases have been reported, Ampligase allows the researcher to use the optimal hybridization temperature and carry out successive cycles of ligation. The \$238 (US) kit consists of 5,000 units of Ampligase, a $10 \times$ reaction buffer, *Sma*I- and *Sall*- digested bacteriophage lambda DNA as a control.

Gel assortment

Boehringer Mannheim has introduced a new **agarose gel** designed for pulsed-field gel electrophoresis (PFGE) applications (Reader Service No. 105). Characteristics of the agarose include high gel strength ($>3,000 \text{ g cm}^{-2}$), low electroendosmosis,



Agarose gel — designed for pulsed-field gel electrophoresis.

and low sulphate concentration (<0.1 per cent), says the company. The high gel strength allows gels to be made with a lower concentration of agarose — 0.6–0.8 per cent instead of 1.2–1.4 per cent. PFGE agarose is available in 10- and 50-g quantities for \$50 and \$175 (US).

As a replacement to acrylamide and agarose gels in some electrophoretic applications, International Biotechnologies is offering Instacryl gel — a **non-toxic, water-soluble copolymer** that is supplied as a pre-mixed stock solution (Reader

Service No. 106). Two forms are available: Instacryl H and L. Instacryl H is recommended for the separation of double-stranded DNA in the 10 to 3,500 base-pair range. Gels are formed in less than 10 minutes, says IBI — just mix with water and buffer and add the cross linker, dithiothreitol. Instacryl L is, on the other hand, recommended for the separation of proteins in the 14 to 330-kDa range. IBI recommends the use of Instacryl H as a 4 per cent stacking gel when using Instacryl L. Instacryl H is available in 250-ml quantities for \$47.50 (US) and Instacryl L in one-litre quantities for \$90 (US).

Non-isotopic labelling

Dako has introduced a new ***in situ* hybridization detection system** for the enzymatic detection of biotinylated nucleic acid probes following hybridization to target DNA sequences in cell or tissue specimens (Reader Service No. 107). The procedure uses streptavidin as the link between the biotinylated probe and the



Dako's *in situ* hybridization kit detects biotinylated probes.

biotinylated reporter molecules of alkaline phosphatase. The alkaline phosphatase/chromogen combination of 5-bromo-4-chloro-3-indolyl phosphate and nitro blue tetrazolium produces an intense blue staining at the site of hybridization. The \$140 (US) kit, which can be used with formalin-fixed paraffin-embedded tissues, contains enough reagents for 200–300 slides, says Dako. Tissues such as liver, kidney, or highly proliferative cells or tissues, which often exhibit high levels of endogenous biotin, may, however, produce a high background signal.

E-Link and E-Link Plus are two new **non-isotopic oligonucleotide labelling kits** from Cambridge Research Biochemicals (Reader Service No. 108). The kits are supplied with sufficient reagents and an easy-to-follow protocol for the covalent labelling of two oligonucleotides with alkaline phosphatase. E-Link Plus includes the chemiluminescent substrate, Lumi-Phos. CRB says results have shown that this combination produces a signal on X-ray film that is equivalent to, or better than, ^{32}P . The E-Link kit does not include the Lumi-Phos substrate. The procedure takes about one hour of 'hands-on time', including a chromatographic step to obtain purified conjugate using the pre-

packed column that comes with the kit. Application notes for Southern and northern blots, dot blots, plaque lifts and *in situ* hybridizations are included in the protocol.

Pure and simple

As a performance upgrade to the Model 340A, Applied Biosystems is launching

ADVERTISEMENTS

SPECIAL PRICE CUSTOM DNA

100 μg , 58.00 FF A BASE
FOR THE FIRST 15 BASES.
49.00 FF, each additional base
BIOPROBE SYSTEMS
CALL : 33-1-48-51-66-22
FAX : 33-1-48-51-59-90

Reader Service No.23

Papers

Bibliographic Database

For an IBM PC/AT compatible computer
£50.00 (inclusive)

Write for your free trial disk to:
DBF Systems (N), PO Box 1374,
London W5 3XX
Please state disk size: 3.5" or 5.25"

Reader Service No.46

EARTHQUAKE

Loma Prieta San Francisco Earthquake, 17th October 1989 . . . a collection of video images.

An authoritative 53 min. documentary of this major seismic event produced by the U.S. Geological Survey.

UK£15.95 or US\$25.00.

VHS format PAL or NTSC versions (PAL sent unless NTSC requested).

MASTERCARD or send cheque to: **MJP**
— **GEOPACKS**, PO Box 23, St. Just,
Penzance, Cornwall, UK. TR19 7JS. Tel.
0736 787808. Fax 0736 787880.

Reader Service No.47

AFRICA AWAITS

SAFARI CAMPING ADVENTURES
1-4 WEEKS FROM £410 - £1,430

TRANS-AFRICA EXPEDITIONS

7-19 WEEKS FROM £1,845 - £2,505
Prices include return flight from London.

Over 50 holidays available in our
colour brochure.



GUERBA
EXPEDITIONS

Dept N 101 Eden Vale Road
WESTBURY Wiltshire BA13 3QX

TEL: 0373 827046 (24 HRS)

Reader Service No.19

ADVERTISEMENTS

Reader Service No.26

Custom DNA

Purified and Shipped in 48 hours,
Worldwide.

100 ug. \$10.00 a base for the first 15 bases,
\$7.50 for each additional base.
Delivered.

Research Genetics

1-800-533-4363

In the United Kingdom 0-800-89-1393
In Canada 1-800-533-4363

FAX 205-536-9016

Reader Service No.29

CUSTOM-MADE & PURIFIED DNA PEPTIDES

\$8/base
50-100 ug
40 hour delivery

Quotes available
50-300 mg
1-3 wks delivery

GUARANTEED QUALITY & QUANTITY

V.G. CORP.

1200 Wonderland Rd. S.
Bldg #7, Unit #2
London, Ontario
Canada N6L 1A8

CALL OR FAX: (519) 652-3758

Reader Service No.50

SINGER
INSTRUMENTS



**COMPLETE SYSTEMS
FOR
XENOPUS OOCYTE INJECTION**

INT. SALES: +44. 984. 40226.

Singer Instrument Co. Ltd.

Roadwater, Watchet, Somerset,
England, TA23 0QL Tel: Washford (0984) 40226.
Telex: 337492 Comcab G

**Teflon[®]
Coated
and
Bare Wire from
MEDWIRE**



Medwire provides
the highest quality
platinum, platinum alloy, silver and stainless
steel wire in both single and multistrand
configurations.

For prompt delivery or additional
information, please contact:

MEDWIRE, an affiliate of Sigmund Cohn Corp.
121 So. Columbus Ave. Mount Vernon, NY 10553
Tel: (914) 664-5300 Fax: (914) 664-5377

*Reg. TM of E.I. DuPont & Co.

the Genepure 341 **nucleic acid purification system**, which is designed to provide faster PCR template production at reduced cost (Reader Service No. 109). Genepure 341 offers a choice of conventional phenol-chloroform or non-organic methods for genomic DNA or RNA purification and can be used with a diverse selection of source materials: whole blood, plants, leukocytes, tissue cell culture and viruses. Applied Biosystems recommends the use of non-organic chemistries for the isolation



Applied Biosystems' nucleic acid purification system — GenePure 341.

of PCR-quality genomic DNA isolated from relatively clean white blood-cell fractions, nuclei preps and cultured cells. The process takes about 1.5 hours, including sample lysis, Proteinase K digestion and alcohol precipitation. The phenol-chloroform method takes about 4 hours and is suitable for purification procedures that involve more complex starting materials, such as whole blood.

Nucleobond AX is a new kit from The Nest Group for the **purification of plasmids** from *Escherichia coli* (Reader Service No. 110). In a four-step procedure, 1-500 ml of *E. coli* cultures can be purified in 3-9 minutes, with no need for HPLC, ultracentrifugation or CsCl gradients, says the company. The kit contains a manual and all the cartridges, buffers, accessories necessary for plasmid purification. Cartridges are disposable, but can be reused if cross contamination is not a concern. The purity of plasmid DNA produced is equal to, or better than, that obtained with CsCl procedures. The resulting plasmid is suitable for sequencing, enzymatic processing, labelling reactions, transformations or transfections, site-directed mutagenesis, and DNA/protein studies.

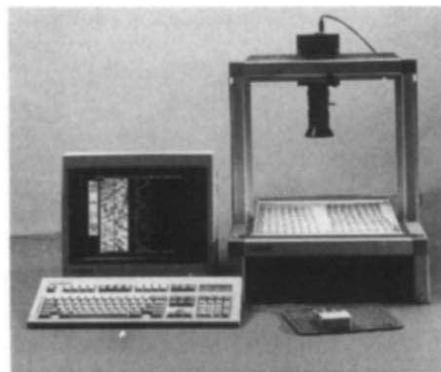
Tools of the trade

Run two minigels simultaneously from a single power source with the latest **dual electrophoresis chamber** from Fotodyne (Reader Service No. 111). Costing \$295 (US), the dual-gel chamber comes complete with built-in power leads, two gel trays, eight combs — four 12-well and four 6-well — and spacers to allow the user to run a single gel if required. The trays have casting gates, so there is no need to tape the ends when casting gels, says Fotodyne. Gels can be viewed *in situ*, as the plexi-

glass bottom of the chamber is UV-transparent.

Additional modules for the Macintosh-based Lasergene **software for molecular biologists** will be available in late Spring, says DNASTAR (Reader Service No. 112). Modules for keyword, pattern and homology searches of GenBank and NBRF-RIR databases, sequence/restriction sites display, translation, comparison and screen editing will cost between \$500 and 3,000 (US). A shotgun sequence assembly program is already available. Alignment programs and protein analysis tools for studying primary and secondary structure of DNA and protein sequences will be available in autumn.

SpeedReader is a **semi-automated DNA sequence film reader** from IntelliGenetics (Reader Service No. 113). The instrument takes input directly from conventional Sanger DNA sequencing autoradiograms and makes electronic adjustments for image sharpness, contrast and electrophoretic abnormalities such as 'smiles' and curving lanes. The pattern of bands is interpreted to produce a sequence file that can be edited and saved in several file formats. With SpeedReader's automatic reading algorithm, an average throughput of two kilobases of edited sequence information per hour can be produced. Identified bands can be marked with a



SpeedReader: DNA sequence reader.

coloured dash so that a quick visual scan can detect mismatched base assignments or unidentified bands. Sequence files are compatible with IntelliGenetics' PC/GENE and Suite, as well as software packages that use EMBL or ASCII formats. The basic system includes a CCD camera, camera stand, light box, digitizer and '286' IBM AT computer and software, and is being offered at an introductory price of \$22,000 (US) for commercial customers. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain additional information about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.